DOCUMENT RESUME

ED 241 973 CS 504 556

AUTHOR Studdert-Kennedy, Michael, Ed.; O'Brien, Nancy, Ed.
TITLE Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, October 1-December 31, 1983.
INSTITUTION Haskins Labs., New Haven, Conn.
SPONS AGENCY National Institutes of Health (DHHS), Bethesda, Md.; National Science Foundation, Washington, D.C.; Office of Naval Research, Washington, D.C.
REPORT NO SR-76(1983)
PUB DATE 83
CONTRACT NICHHD-N01-HD-1-2420; ONR-N00014-83-C-0083
GRANT NICHHD-HD-01994; NICHHD-HD-16591
NOTE 251p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC11 Plus Postage.
DESCRIPTORS *Articulation (Speech); *Auditory Perception; *Communication Research; Consonants; Elementary Education; Language Usage; Linguistic Theory; *Measurement Techniques; Memory; Music; Phonetics; Semantics; Speech Instruction; *Speech Skills; *Vowels

ABSTRACT

One of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical research applications, this report consists of 14 papers. Topics covered in the papers are (1) skilled actions, (2) the control of fundamental frequency declination, (3) selective effects of masking on speech and nonspeech in the duplex perception paradigm, (4) the perception of vowels in consonantal content and in isolation, (5) the relation between children's perceptions of articulation and perceptual adjustment of coarticulary effects, (6) trading relations among acoustic cues in speech perception, (7) the role of release bursts in the perception of (s) stop clusters, (8) changes in spoken Welsh, (9) single format contrast in vowel identification, (10) integration of melody and text in memory for songs, (11) the equation of information and meaning from the perspectives of situation semantics and J. J. Gibson's ecological realism, (12) the equating of information with symbol strings, (13) perception and action, and (14) mapping speech. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SR-76 (1983)

ED241973

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 October - 31 December 1983

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

(The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CS 504 556

Michael Studdert-Kennedy, Editor-in-Chief
Nancy O'Brien, Editor
Margo Carter, Technical Illustrator
Gail Reynolds, Word Processor

ACKNOWLEDGMENTS

The research reported here was made possible
in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994
Grant HD-16591

National Institute of Child Health and Human Development
Contract NO 1-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS-8111470

National Institute of Neurological and Communicative
Disorders and Stroke
Grant NS 13870
Grant NS 13617
Grant NS 18010
Grant NS 07237
Grant NS 07196

Office of Naval Research
Contract N00014-83-C-0083

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President
Franklin S. Cooper,* Assistant to the President
Patrick W. Nye, Vice President
Michael Studdert-Kennedy,* Vice President, Research
Raymond C. Huey,* Treasurer
Bruce Martin, Controller
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Thomas Baer
Patrice Beddor+
Fredericka Bell-Berti*
Shlomo Bentin[1]
Catherine Best*
Gloria J. Borden*
Susan Brady*
Robert Crowder*
Laurie B. Feldman*
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Sarah Hawkins
Satoshi Horiguchi[2]
Leonard Katz*
J. A. Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald P. Shankweiler*
Sigfrid Soli*
Betty Tuller*
Michael T. Turvey*
Douglas Whalen

Technical and Support Staff

Michael Anstett
Margo Carter
Philip Chagnon
Vincent Gulisano
Donald Hailey
Sabina D. Koroluk
Betty J. Myers
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Eric Bateson
Suzanne Boyce
Jo Ann Carlisle
Andre Cooper
Patricia Ditunno
Jan Edwards
Jo Estill
Nancy Fishbein
Carole E. Gelfer
Janette Henderson
Charles Hoequist
Bruce Kay
Noriko Kobayashi
Rena Krakow
Harriet Magen
Sharon Manuel
Richard McGowan
Daniel Recasens
Hyla Rubin
Judith Rubin
Suzanne Smith
Katyanee Svastikula
Louis Tassinary
Ben C. Watson
Deborah Wilkenfeld
David Williams

---

*Part-time
[1]Visiting from Hadassah University Hospital, Jerusalem, Israel
[2]Visiting from University of Tokyo, Japan
+NIH Research Fellow

CONTENTS

I. Manuscripts and Extended Reports

Obituary for Dennis Butler Fry--
Arthur S. Abramson 1-2

Skilled actions: A task dynamic approach--
Elliot Saltzman and J. A. Scott Kelso 3-50

Speculations on the control of fundamental
frequency declination--Carole E. Gelfer,
Katherine S. Harris, Rene Collier,
and Thomas Baer 51-63

Selective effects of masking on speech and
nonspeech in the duplex perception paradigm--
Shlomo Bentin and Virginia A. Mann 65-85

Vowels in consonantal context are perceived more
linguistically than are isolated vowels: Evidence
from an individual differences scaling study--
Brad Rakerd 87-114

Children's perception of [s] and [ʃ]: The
relation between articulation and perceptual
adjustment for coarticulatory effects--
Virginia A. Mann, Harriet M. Sharlin,
and Michael Dorman 115-128

Trading relations among acoustic cues in speech
perception: Speech-specific but not special--
Bruno H. Repp 129-132

The role of release bursts in the perception of
[s]-stop clusters--Bruno H. Repp 133-157

A perceptual analog of change in progress in Welsh--
Suzanne Boyce 159-165

Single formant contrast in vowel identification--
Robert G. Crowder and Bruno H. Repp 167-177

Integration of melody and text in memory for songs--
Mary Louise Serafine, Robert G. Crowder, and
Bruno H. Repp 179-194

The equation of information and meaning from
the perspectives of situation semantics and
Gibson's ecological realism--M. T. Turvey and
Claudia Carello 195-202

A comment on the equating of information with symbol strings--M. T. Turvey and Peter N. Kugler 203-206

An ecological approach to perception and action--M. T. Turvey and Peter N. Kugler 207-236

Mapping speech: More analysis, less synthesis, please--Michael Studdert-Kennedy 237-239

Review (Towards a history of phonetics)--Leigh Lisker 241-243

II. Publications 247

III. Appendix: DTIC and ERIC numbers (SR-21/22 - SR-74) 249-250

I. MANUSCRIPTS AND EXTENDED REPORTS

OBITUARY FOR DENNIS BUTLER FRY*

The death of Professor Dennis B. Fry at the age of 75 on March 21, 1983 was a great blow to his colleagues, his many good friends, his wife Chrystobel, and his three children.

Dennis Fry was born the third of November 1907 in Stockbridge, Hampshire, England. After five years of teaching French, first at Tewkesbury Grammar School and then at Kilburn Grammar School, in 1934 he was appointed Assistant Lecturer in Phonetices at University College London, where he also became Superintendent of the Phonetics Laboratory in 1937. In 1938 he was promoted to Lecturer in Experimental Phonetics. In 1948, the year after the award of his Ph.D. degree, he became Reader in Experimental Phonetics. From 1958 until his retirement in 1975, he was professor of Experimental Phonetics, the first one to hold the title in Britain.

The Department of Phonetics of University College London owed much to Dennis Fry's benevolent yet, I think, firm headship from 1949 to 1971. Indeed, he played an important role in the later absorption of the newly fledged program in linguistics to form the present Department of Phonetics and Linguistics.

Dennis Fry's phonetic interests were broad and included such topics as automatic speech recognition, the perception of lexical stress, children's acquisition of phonology, categorical perception, and the relevance of experimental phonetics for linguistics. He also did much important work on problems of the the deaf, especially deaf children; furthermore, he worked on problems of hearing in aviation during his wartime service (1941-45) with the acoustics laboratory of the RAF Central Medical Establishment. His extensive publications up to the beginning of 1979 are listed in "Essays on the Production and Perception of Speech in Honour of Dennis B. Fry," a special issue of Language and Speech (Vol. 21, Part 4, 1978).

From 1961, the time of the Fourth International Congress of Phonetic Sciences in Helsinki, until his death, Fry served diligently as the President of the Permanent Council for the Organization of International Congresses of Phonetic Sciences. I know that our many excellent congresses regularly gave him pleasure over the successful outcomes of the negotiations that the Council, under his leadership, has been able to carry out with dedicated scholars and scientists in so many places. In his last year he began talking to some of us about encouraging able people in untried parts of the world to mount equally good congresses.

Fry also furthered international cooperation in our field through his link of more than twenty-five years with Haskins Laboratories, first in New York City and then in New Haven. His occasional lengthy visits to do research and his frequent consultations with some of us yielded important results on both sides of the Atlantic Ocean.

---

*Also to appear in Speech Communication.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

In 1958 Dennis Fry founded the journal Language and Speech as an important outlet for broadly interdisciplinary work. He was Editor until 1975 when he persuaded me to join him as Co-Editor; he left the editorship altogether at the end of 1978, three years after his retirement from his professorship.

His talent as a singer, a talent much enjoyed by operatic groups in his region, went with a serious technical interest in music and the singing voice. A very recent example of his publications in that field is his article "The Singer and the Auditorium" in the 1980 volume of the Journal of Sound Vibration.

I should like to end with a personal note. From 1960 on, Dennis Fry's humane and good-humored approach to people and problems gave me a role model that I fear I shall never match. The sudden loss of his warm, caring friendship was hard to take.

Arthur S. Abramson

The University of Connecticut and Haskins Laboratories

Dennis Butler Fry
1907-1983

SKILLED ACTIONS: A TASK DYNAMIC APPROACH

Elliot Saltzman and J. A. Scott Kelso+

Abstract. A task dynamic approach to skilled movements of multidegree of freedom effector systems has been developed in which task-specific, relatively autonomous action units are specified within a functionally defined dynamical framework. Qualitative distinctions among tasks (e.g., the body maintaining a steady vertical posture or the hand reaching to a single spatial target versus cyclic vertical hopping or repetitive hand motion between two spatial targets) are captured by corresponding distinctions among dynamical topologies (e.g., point attractor versus limit cycle dynamics) defined at an abstract task space (or work space) level of description. The approach provides a unified account for several signature properties of skilled actions: trajectory shaping (e.g., hands move along approximately straight lines during unperturbed reaches) and immediate compensation (e.g., spontaneous adjustments occur over an entire effector system if a given part is disturbed en route to a goal). Both of these properties are viewed as implicit consequences of a task's underlying dynamics and, importantly, do not require explicit trajectory plans or replanning procedures. Two versions of task dynamics are derived (control law; network coupling) as possible methods of control and coordination in artificial (robotic, prosthetic) systems, and the network coupling version is explored as a biologically relevant control scheme.

I) Introduction

For animals to function effectively in their environments, their movements must be coordinated in space and time. Though self-evident, this fact

---

+Also University of Connecticut.

Acknowledgment. The preparation of this manuscript was supported in part by Contract No. N00014-83-C-0083 from the U.S. Office of Naval Research, NIH grant NS-13617, and Biomedical Research Support Grant RR-05596. Various aspects of the paper have been formally presented at the Second International Conference on Event Perception, Nashville, Tennessee, June, 1983, and the Engineering Foundation Conference on Biomechanics and Neural Control, Henniker, New Hampshire, July, 1983. We would like to express our appreciation to the following colleagues for their helpful comments on an earlier draft of this paper and for valuable discussions concerning several of the topics therein: James Abbs, John Delatizky, Carol Fowler, Louis Goldstein, Vince Gracco, Neville Hogan, John Hollerbach, Fay Horak, Bruce Kay, Wynne Lee, Rich McGowan, Gin McCollum, Paul Milenkovic, Lewis Nashner, Patrick Nye, Marc Raibert, and Michael Turvey. In addition, we are grateful to Phil Rubin who developed some of the basic software procedures used in the present computer simulations.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

raises a most fundamental issue that has recently attracted a number of disciplines ranging from neuroscience to robotics and cognitive science, viz. how coordination and control arise in complex, multivariable systems. How are the many degrees of freedom adaptively harnessed during coordinated, skilled actions? A deterrent to viable solutions to this problem rests in part in our "limited ability to recognize the significant informational units of movement" (Greene, 1971, p. xviii; see also Szentagothai & Arbib, 1974). For some time, it has seemed questionable to us that nervous systems work through individualized control of component elements, whether they be thought of as joints or muscles. Instead, we believe (and there is an increasing amount of evidence to support the claim) that the many potentially free variables are partitioned naturally into collective functional units within which the component elements may vary relatedly and autonomously. The behavior of these action units or coordinative structures (Fowler, 1977; Kelso, Southard, & Goodman, 1979; Turvey, Shaw, & Mace, 1978) is often exemplified by the existence of relational invariances among kinematic and muscular events during activities as diverse as locomotion, speech, handwriting, and reaching to a target (see Grillner, 1982; Kelso, 1981; Kelso, Tuller, & Harris, 1983; Schmidt, 1982; Viviani & Terzuolo, 1980).

The primary focus of the present paper is to characterize the style of operation of these proposed action units within what we call a task dynamic approach. The term task dynamics follows directly from the view (1) that the degrees of freedom comprising action units are constrained by the particular tasks that animals perform, and (2) that action units are specified in the language of dynamics, not, as is more frequently assumed, in terms of kinematic or muscular variables (cf. Stein, 1982, for an inventory). Thus we propose, and seek to elaborate here, an invariant control structure that is specified dynamically according to task requirements and that gives rise to diverse kinematic consequences.

The paper is organized as follows: First we expand upon those desirable properties of action units that are central to the explication of a task dynamic framework. Second, we present a short tutorial on topological dynamics, a crucial aspect of which is to link the system's geometrical qualities to its dynamics in ways that are task-specific. These steps are precursory to the introduction of the task dynamic approach, two versions of which (control law, network coupling) will be presented. The task dynamic approach will be shown to provide a viable account of such tasks as discrete reaching, bringing a cup to the mouth and turning a handle. It can also offer a principled account of various compensatory behaviors such as those that occur when an arm is perturbed during a reaching movement or when the support base is perturbed during standing. Finally, it will be suggested that the network coupling version of task dynamics both provides an extension of the control law version and offers a new synthesis of recent physiological findings on the planning and control of arm trajectories.

The significance of the task dynamic approach for a theory of coordination and control is that it offers a unified account of certain phenomena that heretofore have required conceptually distinct treatments in the movement literature. In addition, the implications for design and control of robotic and prosthetic devices will be apparent. In fact, the approach shares some but not all of the features of several current developments in manipulator control (cf. Hogan & Cotter, 1982; Raibert, Brown, Chepponis, Hastings, Shreve, & Wimberly, 1981). But before discussing the task dynamic framework

in detail, we will describe the phenomena that led us, in part, to propose the present theoretical approach. Indeed, it is the existence of these phenomena--trajectory shaping and immediate compensation--that constitute the main empirical results that the task dynamic framework is designed to explain.

The first phenomenon, trajectory shaping, refers to the task-specific motion patterns of the terminal devices or end-effectors of the effector systems[1] associated with various types of skill. For example, it has been observed experimentally that, in reaching tasks involving two joints (shoulder and elbow) and two spatial hand motion dimensions, the hands move in quasi-straight-line spatial trajectories from initial to target positions and display single-peaked tangential velocity curves (e.g., Morasso, 1981). Similarly and more obviously, in cup-to-mouth tasks the grasped cup maintains a spillage-preventing, approximately horizontal orientation en route from table to mouth.

The second phenomenon, immediate compensation, refers to the fact that skilled movements show task-specific flexibility in attaining the task goal. If one part of the system is perturbed, blocked, or damaged, the system is able to compensate (assuming the disturbance is not "too big") by reorganizing the activities of the remaining parts in order to achieve the original goal. Further, such readjustments appear to occur automatically without the need to detect the disturbance explicitly, replan a new movement, and execute the new movement plan. Kelso, Tuller, and Fowler (1982) have demonstrated such behavior in the speech articulators (jaw, upper and lower lip, tongue body) when subjects produced the utterances /baeb/ or /baez/ across a series of trials in which the jaw was occasionally and unpredictably tugged downward while moving upward to the final /b/ or /z/ closure (see also Abbs & Gracco, in press; Folkins & Abbs, 1975). The system's response to the jaw perturbation was measured by observing the motions of the jaw and upper and lower lips as well as the electromyographic (EMG) activities of the orbicularis oris superior (upper lip), orbicularis oris inferior (lower lip), and genioglossus (tongue body) muscles. The investigators found relatively "immediate" task-specific compensation (i.e., 20-30 ms from onset of jaw pull to onset of compensatory response) in remote articulators to jaw perturbation. For /baeb/ (in which final lip closure is crucial) they found increased upper lip activity (motion and EMG) relative to the unperturbed control trials but normal tongue activity; for /baez/ (in which final tongue-palate constriction is important) they found increased tongue activity relative to controls, but normal upper lip motion. The speed of these task-specific patterns indicates that compensation does not occur according to traditionally defined "intentional" reaction time processes, but rather according to an automatic, "reflexive" type of organization. However, such an organization is not defined in a hard-wired input/output manner. Instead, these data imply the existence of a selective pattern of coupling or gating among the component articulators that is specific to the utterance produced. Essentially, then, such compensatory behavior represents the classic phenomenon of motor equivalence (Hebb, 1949; Lashley, 1930) according to which a system will find alternate routes to a given goal if an initially traversed route is unexpectedly blocked.

What type of sensorimotor organization could generate, in a task-specific manner, both characteristic-trajectory patterns for unperturbed movements and spontaneous, compensatory behaviors for perturbed movements? We believe that a task-dynamic approach provides at least the beginnings of a cohesive answer to this question. Let us examine these issues, then, beginning with an overview of action unit properties.

II) Units of Action

There are three major points to be made concerning our description of action units:

1. Functional definition; Special purpose device. Action units are defined abstractly in a functional, task-specific fashion and span an ensemble of many muscles or joints. Thus, they are not defined in a traditional reductionist sense relative to single muscles and/or joints, nor are they hard-wired input-output reflex arrangements. These units serve to constrain the muscle/joint components of the collective to act cooperatively in a manner specific to the task at hand. For different skilled actions, performers transform the limbs temporarily into different special purpose devices whose functions match the tasks being performed. Thus, an arm can become a retriever, puncher, or polisher; a leg may become a walker or kicker; the body can become a dancer or swimmer; the speech organs may become talkers, singers, chewers, or swallowers, etc.

2. Autonomy. Action units operate relatively autonomously and are to a large extent self-regulating. That is, once a given functional organization is established over a muscle/joint collective, the system achieves its goal with minimal "voluntary" intervention. In later discussions of the mathematics of task dynamics, we will also indicate that action units are relatively autonomous in a strict mathematical sense, i.e., the equations describing task-dynamic systems are not explicit functions of an independent time variable.

3. Dynamics. Action units are defined in the language of dynamics, not kinematics (e.g., Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980). The behavior of an effector system is controlled by a task-specific patterning of the system's dynamic parameters (e.g., stiffness, damping, etc.) according to the abstract functional demands of the performed skill. Such dynamical patterning serves to convert the effector system into the appropriate task-demanded special purpose device. Further, this patterning both generates the observable motions that are characteristic of that skill and underlies the ability to compensate spontaneously for unpredicted disturbances. There is no explicit plan for the desired kinematic trajectory in the action unit, nor is there an explicit contingency table of replanning procedures for dealing with unexpected perturbations. Rather, task-specific kinematic trajectories and compensatory behaviors emerge from, or are implicit consequences of, the action unit's dynamics. In this sense, most robots (with at least one notable exception, i.e., Raibert et al., 1981) have no skills, but are controlled instead as general-purpose devices using the same dynamical structure for all types of tasks, e.g., spatial trajectory planning for the terminal device, conversion to a joint velocity plan, and joint velocity servoing for both manipulators (e.g., Whitney, 1972) and hexapod walker legs (e.g., McGhee & Iswandhi, 1979).

Given the above three points, one can formulate the problem of skill learning as that of designing an action unit or coordinative structure whose underlying dynamics are appropriate to the skill being learned. That is, in acquiring a skill one is establishing a one-to-one correspondence between the functional characteristics of the skill and the dynamical pattern underlying the performance of that skill. This correspondence between dynamics and function is perhaps the key concept underlying the task-dynamic approach. To explore it more fully we will now: a) examine the geometric notion of topology

as it relates to a system's dynamics; and b) describe how functionally specific dynamical topologies can be used to specify task-specific action units or coordinative structures.

III) Topology and Dynamics

Quite (and perhaps too) simply in the context of skilled action, topology refers to the qualitative aspects of a system's dynamics, e.g., whether a system's dynamics generate 1) a discrete motion to a single target or 2) a sustained cyclic motion between two targets. For a one-degree-of-freedom rotational system such as the elbow joint (flexion-extension degree of freedom) the first motion type might correspond to a positioning task with a single joint angle target, while the second might correspond to a reciprocal tapping task between two joint angle targets. What sort of dynamics might underlie these qualitatively different tasks? For the discrete task, several investigators have hypothesized that the system can be modeled as a damped mass-spring system (e.g., Cooke, 1980; Fel'dman, 1966; Kelso, 1977; Kelso & Holt, 1980; Polit & Bizzi, 1978; Schmidt & McGown, 1980). Such a dynamical system may be described by the following equation of motion:

$$I\ddot{x} + b\dot{x} + k(x - x_0)=0, \text{ where} \qquad (1)$$

I = moment of inertia about the rotation axis;
b = damping (friction) coefficient;
k = stiffness coefficient;
x_0 = equilibrium angle;
x, $\dot{x}$, $\ddot{x}$ = angular displacement and its respective first and second time derivatives.

If we assume a set of constant dynamical parameters (I, b, k, x_0), then the behavior of this system can be characterized by its point stability or equifinality, in that it will come to rest at the specified x_0 "target" despite various initial conditions for x and $\dot{x}$ and despite any transient perturbations encountered en route to the target.

The behavior of such systems can be displayed graphically in two different ways. In Figure 1A, the angle of an underdamped mass-spring with constant coefficients is plotted as a function of time for a given set of initial conditions and with no perturbations introduced. Defining the equilibrium or rest angle as aligned with the abscissa, one observes the system's point stability in the progressive decay of the amplitude to the steady state rest angle. In Figure 1B, the same trajectory is represented alternatively in the phase plane for which the abscissa and ordinate correspond to x and $\dot{x}$, respectively, and in which the system's x_0 is located at the phase plane origin. In the phase plane, the system's point stability may be observed as the trajectory spirals down to the origin. Theoretically, if one were to plot the phase plane trajectories corresponding to all possible initial conditions, one would fill the plane with qualitatively similar decaying trajectories defining, thereby, the system's phase portrait. The qualitative "shape" of the system's phase portrait reflects the system's dynamical topology, i.e., the characteristic relations among the system's underlying dynamic parameters. For the type of system described by equation 1, the corresponding phase portrait represents the topology of a point attractor (Abraham & Shaw, 1982), and the

Figure 1. Point attractor system. A. position vs. time; B. Velocity vs. position (phase portrait).

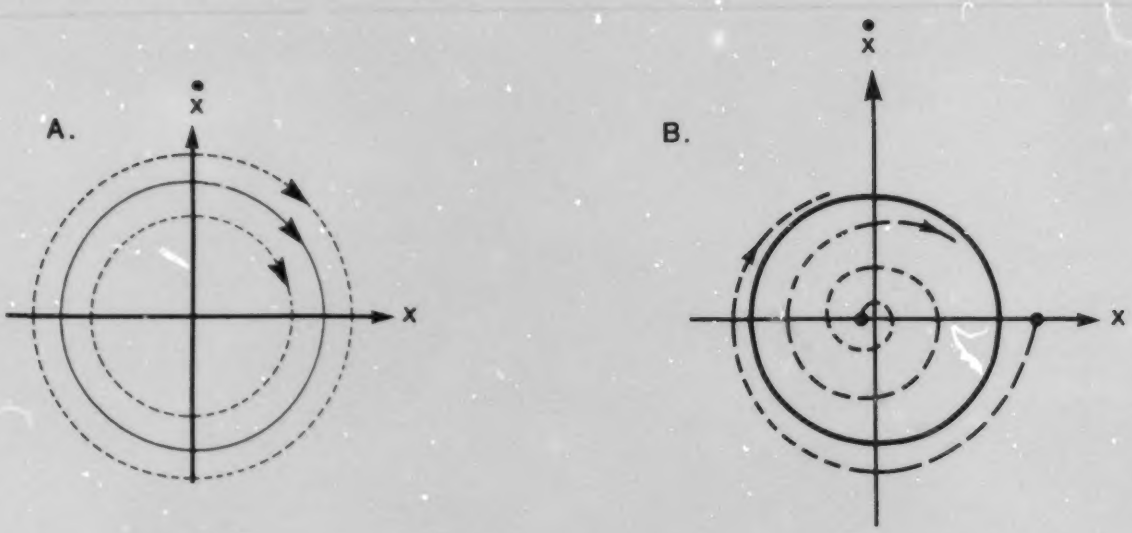


Figure 2. Phase portraits for neutrally stable system (A) and periodic attractor system (B), showing system trajectories for several initial conditions.

underlying dynamics may be described as point attractor dynamics. As a model of discrete positioning tasks, such point attractor dynamics are appealing since the same underlying topology will accommodate different trajectory characteristics (e.g, peak velocity, movement time) and target positions by specification of different values for the system's dynamic parameters.

Obviously, another type of dynamics is required to generate the kinematics observed in a sustained cyclic elbow (or finger, e.g., Kelso, Holt, Rubin, & Kugler, 1981) rotation between two target angles. Perhaps the simplest dynamic scheme corresponds to that of an undamped mass-spring system or harmonic oscillator, with the following equation of motion:

$$I\ddot{x} + k(x - x_o) = 0, \qquad (2)$$

where all symbols are defined as in equation (1). The solid line trajectory in Figure 2A represents the phase plane orbit of such a system, which oscillates about the origin (x_0) with an amplitude that is determined by the system's total mechanical energy, and whose angular targets correspond to the system's maximum and minimum angular limits. However, this type of system does not provide a satisfactory model for the cyclic elbow task for two reasons: a) it represents the ideal frictionless case and no real world system is frictionless. Adding friction to equation (2) would simply convert it to equation (1), leaving a point-attractor dynamics unsuitable for any sustained cyclic task; and b) the system described by equation 2 is only <u>neutrally stable</u> in that the oscillation amplitude is extremely plastic with respect to both initial system energy (determined by initial conditions of position and velocity) and transient changes (perturbations) in energy imposed during oscillations. For example, the dotted trajectories in Figure 2A represent oscillations of the same system as does the solid trajectory. However, the inner and outer dotted orbits show the oscillations corresponding to smaller and larger amplitude initial conditions, respectively, relative to the solid orbit. Clearly then, for a task whose oscillation amplitude is crucial, a neutrally stable system is undesirable.

One can overcome the above shortcomings of an undamped mass-spring dynamics, however, by moving to an alternate <u>periodic attractor</u> (Abraham & Shaw, 1982) dynamical model, with the following equation of motion:

$$I\ddot{x} + b\dot{x} + k(x - x_o) = f(x,\dot{x}), \text{ where} \qquad (3)$$

$I, b, k, x_o, x, \dot{x}, \ddot{x}$ are as in equations 1 and 2; and
$f(x,\dot{x})$ = nonlinear escapement function of the system's current $x, \dot{x}$.

This system's behavior is characterized by the three phase plane trajectories seen in Figure 2B corresponding to three different sets of initial conditions. The solid trajectory represents a motion starting at either target, and the inner and outer dotted trajectories represent motions starting inside and outside, respectively, of the target-to-target angular range. It can be seen that these trajectories converge onto the solid orbit, which is described as a <u>stable limit cycle</u> or periodic attractor. In fact, all trajectories (except those starting exactly at x_o) converge to the limit cycle, and the corresponding phase portrait captures the topology of this periodic attractor

dynamical system. The reason for this orbital stability lies in the nature of the nonlinear escapement term, $f(x,\dot{x})$, seen in equation 3.[2] Basically this term is the means by which the system taps an external energy source in a self-gated manner, i.e., energy is gated in or out of the system as a function of the system's current $x,\dot{x}$ state. On the limit cycle, the energy tapped per cycle from the external reservoir is equal to the energy dissipated per cycle both by the system's intrinsic damping properties (i.e., $b\dot{x}$) and the system's escapement function. Inside the limit cycle, the energy tapped per cycle is greater than that dissipated, and trajectories grow or spiral out to the limit cycle; outside the limit cycle, the converse situation holds, and trajectories decay or spiral down to the limit cycle (cf. Minorsky, 1962).

The above examples illustrate how particular distinct task functions (discrete positioning vs. cyclic alternation) may be modeled by topologically distinct dynamical systems. It should be noted, however, that both tasks and dynamics were defined in single degree of freedom systems. In these cases one dimensional motions were demanded by the tasks and these task requirements were mapped directly onto corresponding dynamical control types at a single joint. This style of control, in which task-specific sets of constant dynamic parameters are defined with respect to control at single joints or articulator degrees of freedom, may be labeled articulator dynamics. Real world tasks seldom involve such simple one-to-one mappings of task demands into sets of constant articulator dynamic parameters. Consider, for example, the two dimensional discrete reaching task discussed earlier in the Introduction involving two articulator degrees of freedom (shoulder, elbow) and two spatial dimensions of terminal device (hand) motion. Extending an articulator dynamics approach to this more complex task meets with only limited success, providing a reasonable account of final position control but failing to account for the observed characteristic quasi-straight line hand trajectory patterns (Delatizky, 1982). More specifically, in this two dimensional task the arm is effectively nonredundant[3] (e.g., Saltzman, 1979) and, given the anatomical limits on joint angular excursion, there is a unique mapping from hand position to arm configuration (i.e., the set of shoulder and elbow angles; arm posture). Therefore, if one defines constant point attractor dynamics at each joint with rest angles corresponding to the target arm configuration (and thus target hand position), the hand/arm will exhibit equifinality by attaining the desired target position/configuration despite variation in initial position/posture and despite transient disturbances encountered en route to the target. However, as mentioned above (and to be explained in greater detail below), such an articulator dynamics approach fails to account for the characteristic trajectory patterns seen in these reaching tasks, i.e., this approach does not "favor straight line movements over other movements" (Hollerbach, 1982, p. 190).

At this point, then, those committed to a dynamical account of coordinated movement face a nasty dilemma. The conceptually parsimonious account of motor control via articulator dynamics no longer appears valid. That is, the elegance of the articulator dynamic account for single degree of freedom tasks lay in its use of a set of constant, task specific, articulator-dynamic parameters to generate a potentially infinite number of task-appropriate kinematic trajectories. The failure of such an approach when extended to trajectory shaping in a multidegree-of-freedom task as simple as reaching shows that searching for invariant task-specific action units at the level of articulator dynamics is likely to be a frustrating and probably pointless endeavor. What type of principles or control structures might underlie the trajectory con-

straints on arm motion during reaching tasks? There are (at least) two alternative accounts. The first is simply to abandon the dynamical approach altogether, and invoke explicit kinematic trajectory plans as sources for the characteristic constraints on motion patterns observed in different tasks. Such an approach has been generally adopted in the field of robotics (e.g., Hollerbach, 1982; Saltzman, 1979), and has been described in the following fashion by Hollerbach (1982):

> A hierarchal movement plan is developed at three levels of abstraction...The top level is the object level, where a task command, such as 'pick up the cup', is converted into a planned trajectory [italics added] for the hand or for the object held by the hand. At the joint level the object trajectory is converted to co-ordinated control of the multiple joints of the human or robotic arm. At the actuator level the joint movements are converted to appropriate motor or muscle activations.

Alternatively, the second account involves defining dynamical control topologies at a level of task description more abstract than the level of individual joints. This leads us to a task-dynamic account of skilled actions.

IV) Task Dynamics

Previous articulator-dynamic descriptions of skilled movement provided plausible accounts of only a very limited type of data: that obtained in laboratory tasks where uni-dimensional tasks mapped directly onto control at a single joint. For example, a discrete target acquisition task was thought to involve specifying the dynamic rest angle parameter corresponding to the task's target joint angle. However, given the failure of articulator dynamics to account for data observed in more complex multivariable tasks, one begins to suspect that this approach might be inappropriate even as a model for control of single variable tasks. More specifically, one reaches the conclusion that the dynamics underlying control of a single joint task might be defined more abstractly than at the articulator level[4] (or joint level; see Hollerbach's, 1982, quote above).

On the basis of a logical analysis of performances across a set of multivariable real world tasks, two common aspects shared by all tasks become evident: a) tasks are typically defined for the terminal devices associated with task-relevant multidegree-of-freedom effector systems (e.g., the grasped cup and arm-trunk, respectively, for a cup-to-mouth task); and b) tasks typically demand characteristic patterns of motion or force[5] by these terminal devices relative to a set of task-specific spatial axes or degrees of freedom. Thus, a given task type can be associated with a corresponding task-spatial coordinate system (task space) that is defined on the basis of both the terminal devices and the environmental objects or surfaces relevant to the task's performance. In fact, Soechting (1982) has presented evidence from a pointing task involving the elbow joint that implies that the controlled variable is not joint angle per se, but rather the orientation angle of the forearm in a spatial coordinate system defined relative to an environmental reference (e.g., the floor surface, or gravity vector, etc.) or the actor's trunk. This suggests that a task-spatial coordinate system might indeed be the appropriate level at which to characterize a skilled action.

The central tenet of the task-dynamic approach is that a set of constant task-dynamic parameters can be defined for each of a given skill's task-space degrees of freedom, defining, thereby a one-to-one correspondence between the functional characteristics of the skill and the task-dynamical topology underlying that skill's performance. In other words, skill-invariant action units are defined functionally relative to a given skill's task space and underlying task-spatial dynamics (more simply, task dynamics). Such sets of constant task-dynamical parameters may be used to define changing patterns of articulator-level dynamic parameters (e.g., joint stiffnesses, dampings, rest angles, etc.) according to two related versions (control law and network coupling) of the task dynamic approach. The evolving constraints on articulator dynamics serve to convert a given skill's effector system into an appropriate special purpose device whose individual components (i.e., articulator degrees of freedom) act cooperatively in a manner specific to the task at hand. It should be remembered for purposes of comparison that the articulator-dynamics approach postulated sets of constant dynamic parameters at the individual joint level; a given set would underlie the resultant variety of equifinal kinematic trajectories for a given type of single degree of freedom task. In contrast, the task dynamic approach postulates sets of constant dynamic parameters at the task-spatial level for given types of multivariable tasks. A given set of such parameters would: a) underlie directly an articulator-state dependent patterning of articulator dynamic parameters; and b) underlie ultimately the task-specific trajectory patterns and compensatory behaviors observed during task performances. We will now provide an overview of the specifics of the task-dynamic approach, using a relatively simple arm reaching task for illustrative purposes. A schematic of the approach and the coordinate transformations involved is shown in Figure 3.

A. Task dynamics; Task network

1. Task-space. A task-dynamic approach to a given skill begins with an abstract, functional description of that skill's task space. Such a description has three parts. First, the relevant terminal devices and goal objects or surfaces are defined. Second, an appropriate number of task axes or degrees of freedom are defined relative to the terminal device and goal referents; and finally, an appropriate type of task dynamic topology is defined along each task axis. For a discrete reaching task in two spatial dimensions, the corresponding task space is modeled as a two-dimensional point attractor and is illustrated in Figure 4A. In this figure, the reach target (x) defines the origin of a t_1t_2 Cartesian coordinate system. Axis t_1 (the "reach axis") is oriented along the line from the target to the initial position of the terminal device (open circle), which is modeled as an abstract point task-mass. Axis t_2 is defined orthogonal to t_1 and measures deviations of the task mass from the reach axis. The task-mass is allowed to assume any t_1t_2 position (filled circle) during task performance, and may be considered an abstract point mass since it is not tied to any particular effector system. The equations of motion corresponding to axes t_1 and t_2 are as follows:

$$m_T\ddot{t}_1 + b_{T1}\dot{t}_1 + k_{T1}t_1 = 0$$
$$m_T\ddot{t}_2 + b_{T2}\dot{t}_2 + k_{T2}t_2 = 0 \qquad (4)$$

where,

m_T = task-mass coefficient;

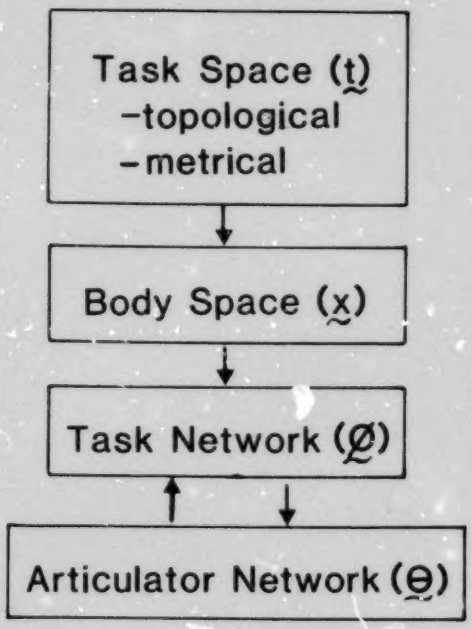


Figure 3. Overview of descriptive levels in task dynamic approach.

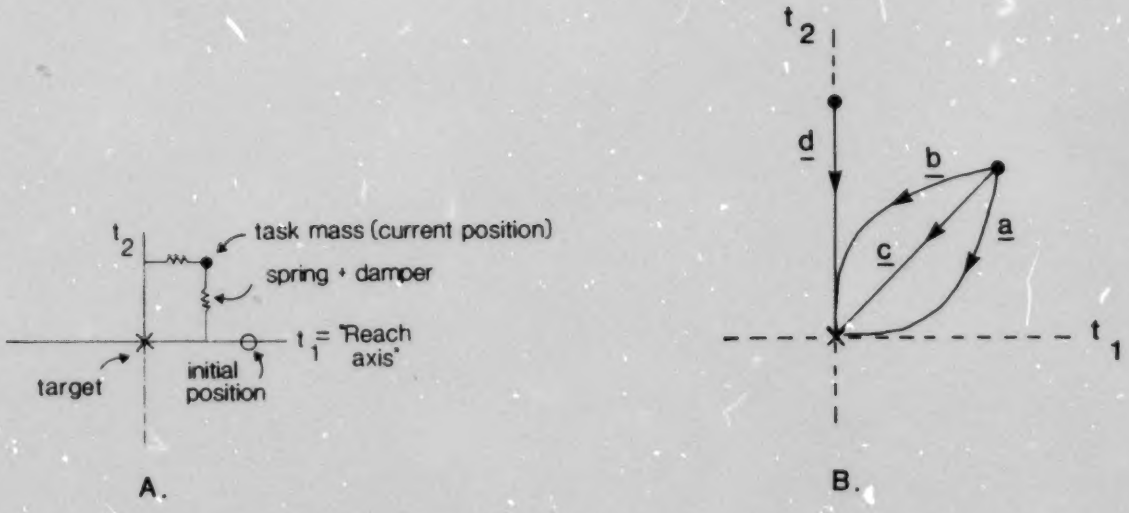


Figure 4. A. Discrete reaching (task space); B. System trajectories corresponding to different task axis weightings and initial conditions.

b_{T1}, b_{T2} = damping coefficients;
k_{T1}, k_{T2} = stiffness coefficients.

In Figure 4A the corresponding damping and stiffness elements are represented in lumped form by the squiggles in the lines connecting the task mass to axes t_1 and t_2. Equation (4) describes a linear, uncoupled set of task-spatial dynamic equations, whose terms are defined in units of force, and whose dynamic parameters are constant. This equation can be represented in matrix form as:

$$M_T\ddot{\underset{\sim}{t}} + B_T\dot{\underset{\sim}{t}} + K_T\underset{\sim}{t} = 0 \text{ , where} \qquad (5)$$

$$M_T = \begin{bmatrix} m_T & 0 \\ 0 & m_T \end{bmatrix}; \quad B_T = \begin{bmatrix} b_{T1} & 0 \\ 0 & b_{T2} \end{bmatrix};$$

$$K_T = \begin{bmatrix} k_{T1} & 0 \\ 0 & k_{T2} \end{bmatrix}$$

It should be noted that there are two nested structures of dynamical constraints at the task space level. The first constraint structure is defined globally, and serves to establish a task-specific dynamical topology. In our reaching example, these global constraints on the task-dynamic coefficients specified point attractor topologies along each task axis. Additionally, however, a set of locally defined, metrical constraints serve to tune the task-spatial dynamic parameters (M_T, B_T, K_T) according to current task demands. Thus, in the reaching example m_T designates the perceptually estimated mass of the terminal device (i.e., gripper + any grasped object-to-be-moved), and B_T and K_T are specified, for example, according to the desired or required damping ratios ($\zeta_{Ti} = b_{Ti}/[2\sqrt{m_T k_{Ti}}]$; i = 1,2) and settling times ($T_{si} = 4/[\zeta_{Ti}\sqrt{k_{Ti}/m_T}]$; i = 1,2; i.e., the time required for the system to settle within 2% of the target amplitude; Dorf, 1974) along each task axis.

The movements of the task mass in reaching space display two properties highly desirable for the terminal devices of real world reaching tasks. Due to the point attractor dynamics, the movements will exhibit equifinality in that the task mass will come to rest at the target regardless of initial position (i.e., by definition, initial distance along t_1) and velocity (i.e., initial direction and speed of task space motion) and despite transient perturbations introduced en route to the target. Additionally, the task mass will show straight line trajectories during unperturbed motions to the target, since in this case the system is effectively one-dimensional by virtue of the definition of the reach axis. However, motions in which the task mass is perturbed away from the reach axis will display trajectory shapes that depend on the relative values of k_{T1} and k_{T2} (assuming equivalent damping properties along each axis) as well as the position in t_1-t_2 space where the perturbation "deposits" the task mass (see Figure 4B). Assuming critical damping along both task axes (i.e., $\zeta_{Ti} = 1.0$; i = 1,2) and a post-perturbation velocity of zero, then: a) when $k_{T1} < k_{T2}$, the task mass will approach the reach axis faster than axis t_2; b) when $k_{T1} > k_{T2}$, the task mass will approach axis t_2 faster than t_1; and c) when $k_{T1} = k_{T2}$, the task

mass will approach t_1 and t_2 at the same speed, showing a straight line post-perturbation trajectory to the target. A straight line post-perturbation trajectory will also result if, regardless of the relative values of k_{T1} and k_{T2}, the task-mass is deposited precisely on the t_2 axis (Figure 4B, trajectory d). The reason for these relationships between perturbed position, relative axis stiffness, and trajectory shape lies in the shape of the potential energy functions corresponding to these different relative k_{T1} and k_{T2} values, and the resultant constraints placed on the ensuing motions of the task-mass when starting at various post-perturbation locations on these manifolds (see Hogan, 1980, for a more detailed discussion of potential energy functions and spring stiffnesses in a similar two-dimensional mass-spring system). Finally, note that these free and perturbed trajectories evolve as implicit consequences of the underlying task space dynamics and, therefore, do not reflect the use of either explicit trajectory plans or replanning procedures.

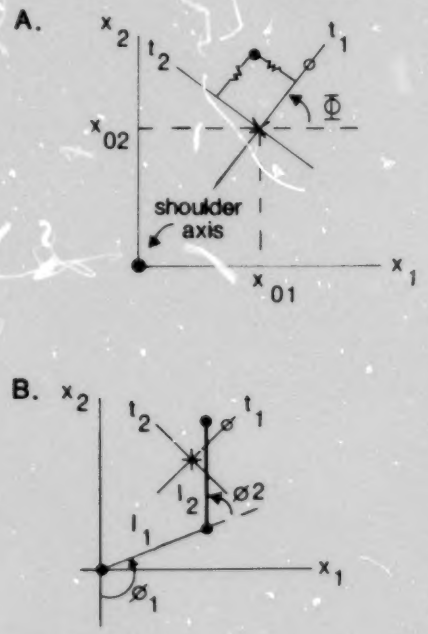


Figure 5. Discrete reaching: A. Body space. Task space is embedded in a shoulder-centered coordinate system; B. Task network. Body space description is transformed into joint variable form of massless model arm.

2. Body Space. The above patterns of task spatial dynamic parameters were defined relative to an environmentally defined goal location and an abstract disembodied terminal device. If these patterns are to be useful to a performer, they must first be transformed into egocentric or body spatial form (e.g., Saltzman, 1979). Such a transformation must be sensitive to the current spatial or geometric relationship between the performer and the task space. As illustrated in Figure 5A for a reaching task, this corresponds to locating and orienting the task space relative to a body spatial (x_1, x_2) coordinate system whose origin corresponds to the current location of the

shoulder's rotation axis. Thus, the terminal device's (task-mass's) current location may be specified in x_1x_2 coordinates. Further, the set of locally defined constraints given by the spatial relationship between task and body spaces serve to tune the body spatial dynamic parameters $\underset{\sim}{x}_0 = (x_{01}, x_{02})^T$ (the location of the task space origin in body space coordinates) and Φ (the orientation angle between the task space's reach axis t_1 and body space axis x_1). Given this information, the task-spatial dynamical pattern may be transformed into a corresponding body or shoulder spatial pattern. The resulting set of linear body-spatial equations of motion for the task's terminal device are defined in matrix form as follows (Note: In these and the following equations, a superscript T denotes the vector matrix transpose operation):

$$M_B\ddot{\underset{\sim}{x}} + B_B\dot{\underset{\sim}{x}} + K_B\Delta\underset{\sim}{x} = \underset{\sim}{0}, \quad \text{where} \qquad (6)$$

$M_B = M_TR$, where M_T = task space mass matrix; and

R = the rotation transformation matrix with elements r_{ij} converting task space variables into body space form;

$$= \begin{bmatrix} \cos\Phi & \sin\Phi \\ -\sin\Phi & \cos\Phi \end{bmatrix} ;$$

$B_B = B_TR$, where B_T = task space damping matrix

$K_B = K_TR$, where K_T = task space stiffness matrix

$\Delta\underset{\sim}{x} = \underset{\sim}{x} - \underset{\sim}{x}_0$, where $\underset{\sim}{x} = (x_1, x_2)^T$, the current body space position vector of terminal device; and

$\underset{\sim}{x}_0 = (x_{01}, x_{02})^T$, the body space position vector of the task space origin.

One should note that equation (6), unlike equations (4) and (5), represents a set of (usually) coupled, autonomous body spatial dynamic equations (i.e., the off-diagonal terms are generally non-zero) due to the rotation transformation. However, as in the case of the task-dynamic parameters, the terms of (6) are defined in force units and the resultant set of body spatial dynamic parameters is constant.

3. Joint variables; Task Dynamic Network. The above patterns of body spatial dynamic parameters were defined with reference to motions of an abstract terminal device disembodied from its effector system. These patterns may be further transformed into an equivalent expression based on the joint-variables of a massless "model" effector system. Like the transformation from task-space to body space, this transformation is a strictly kinematic one and involves only the substitution of variables defined in one coordinate system for variables defined in another coordinate system. As illustrated in Figure 5B, this corresponds to expressing body spatial variables ($\underset{\sim}{x}$, $\dot{\underset{\sim}{x}}$, $\ddot{\underset{\sim}{x}}$) as functions of an arm model's kinematic variables ($\underset{\sim}{\phi}$, $\dot{\underset{\sim}{\phi}}$, $\ddot{\underset{\sim}{\phi}}$), where $\underset{\sim}{\phi} = (\phi_1, \phi_2)^T$, ϕ_1 = shoulder angle defined relative to axis x_2, ϕ_2 = elbow angle defined relative to the upper arm segment. It should be emphasized that the model arm used for this transformation is defined in

kinematic terms only (i.e., the proximal and distal segments have lengths l_1 and l_2, respectively, but no masses), and that the arm's proximal (shoulder) and distal (wrist) ends are attached to the body space origin and the terminal device/task mass, respectively. The transformed equation is as follows (see Appendix A for details):

$$M_B J\ddot{\underset{\sim}{\phi}} + B_B J\dot{\underset{\sim}{\phi}} + K_B \Delta\underset{\sim}{x}(\underset{\sim}{\phi}) = -M_B V\dot{\underset{\sim}{\phi}}_p, \text{ where} \qquad (7)$$

M_B, B_B, K_B are the same constant matrices used in equation (6);

$\Delta\underset{\sim}{x}(\underset{\sim}{\phi}) = \underset{\sim}{x}(\underset{\sim}{\phi}) - \underset{\sim}{x}_o$, where

$\underset{\sim}{x}(\underset{\sim}{\phi}) = (x_1(\underset{\sim}{\phi}), x_2(\underset{\sim}{\phi}))^T$, the current body space position vector of the terminal device expressed as a function of current joint angles;

$\underset{\sim}{x}_0$ = the same constant vector used in equation (6);

$J = J(\underset{\sim}{\phi})$, the Jacobian transformation matrix whose elements J_{ij} are partial derivatives, $\partial x_i / \partial \phi_j$, evaluated at the current $\underset{\sim}{\phi}$;

$\dot{\underset{\sim}{\phi}}_p = (\dot{\phi}_1^2, \dot{\phi}_1\dot{\phi}_2, \dot{\phi}_2^2)^T$, the current joint velocity product vector; and

$V = V(\underset{\sim}{\phi})$, a matrix of coefficients associated with $\dot{\underset{\sim}{\phi}}_p$ introduced during the kinematic transformation and evaluated at the current $\underset{\sim}{\phi}$.

One should note that the matrix products in equation (7) are not constant, but are nonlinearly dependent on the current arm model posture $\underset{\sim}{\phi}$ via the configuration dependence of the $J(\underset{\sim}{\phi})$ and $V(\underset{\sim}{\phi})$ matrices. Further, although equation (7) is expressed in terms of articulator or effector system variables, it is by no means an articulator-dynamic equation. Rather, it is simply the body-spatial dynamic equation (6) rewritten in the articulator-kinematic variables of a massless arm model with no reference to the actual mechanics of a performer's corresponding real arm. Its terms, in fact, are still defined in units of force not torque. Thus, if the initial state $(\underset{\sim}{\phi}_I, \dot{\underset{\sim}{\phi}}_I)$ for the arm model in equation (7) specifies an initial body-spatial wrist position and velocity equal to the initial position and velocity for the task-mass in equation (6), the arm model's joints will change (via equation (7)) in such a way that the wrist moves along exactly the same trajectory as would the abstract terminal device (via equation (6)).

Equation (7) may be rewritten in units of angular acceleration:

$$\ddot{\underset{\sim}{\phi}} + J^{-1} M_B^{-1} B_B J\dot{\underset{\sim}{\phi}} + J^{-1} M_B^{-1} K_B \Delta\underset{\sim}{x}(\underset{\sim}{\phi}) + J^{-1} V\dot{\underset{\sim}{\phi}}_p = \underset{\sim}{0} \qquad (8)$$

For reasons to be elaborated further in the sections to follow, we consider equation (7) to define the task dynamic network (task network) for our reaching task example since, in effect, this equation describes a network of task- and context-specific dynamical relations among the arm model's articulator-kinematic variables. Ultimately, however, a reaching task is performed by

a real arm whose motions and responses to perturbations are shaped according to task-specific, evolving patterns of articulator-dynamic parameters. In the task dynamic approach, constraints are supplied for these articulator dynamics with reference to the task network equation (8).

We will now review the basic articulator-dynamics of a simple two-jointed arm, and then discuss two alternative ways in which equation (8) might be used to constrain these dynamics for a reaching task.

B. Articulator dynamics; Articulator network

For the purpose of simplicity, we will restrict our discussion to a two-joint, two-segment effector system whose segments ("upper arm" and "forearm") have lengths l_1 and l_2, with masses m_1 and m_2 uniformly distributed along the respective segment lengths. Assuming frictionless revolute joints (θ_1, θ_2; defined in the same manner as for the model arm) and no gravity, the passive mechanical (no controls) articulator dynamic equations of motion, whose terms are defined in units of torque, are (see Appendix B for details):

$$M_A\ddot{\underset{\sim}{\theta}} + S_A\dot{\underset{\sim}{\theta}}_p = \underset{\sim}{0} \quad \text{, where} \qquad (9)$$

$M_A = M_A(\underset{\sim}{\theta})$, the 2 X 2 acceleration sensitivity matrix associated with inertial torques, whose elements are functions of the current linkage configuration, $\underset{\sim}{\theta}$. The subscript "A" denotes articulator dynamic elements;

$S_A = S_A(\underset{\sim}{\theta})$, a 2 X 3 matrix associated with coriolis torques (related to joint velocity cross products) and centripetal torques (related to squares of joint velocities), whose elements are functions of the current linkage configuration, $\underset{\sim}{\theta}$.

With controls included, this equation becomes:

$$M_A\ddot{\underset{\sim}{\theta}} + S_A\dot{\underset{\sim}{\theta}}_p + B_A\dot{\underset{\sim}{\theta}} + \underset{\sim}{\tau}_{As} + \underset{\sim}{\tau}_{Aa} = \underset{\sim}{0}, \text{ where} \qquad (10)$$

or

$$K_A\underset{\sim}{\Delta\theta}$$

B_A = a 2 X 2 control damping matrix;

$\underset{\sim}{\tau}_{As}$ = a 2 X 1 control spatial-spring torque vector;

K_A = a control 2 X 2 joint-stiffness matrix;

$\underset{\sim}{\Delta\theta} = \underset{\sim}{\theta} - \underset{\sim}{\theta}_o$, where $\underset{\sim}{\theta}_o$ = a 2 X 1 control reference configuration vector; and

$\underset{\sim}{\tau}_{Aa}$ = a 2 X 1 control additional torque vector, whose function will be described more fully in the following section on Control Laws.

Equation (10) may be rewritten as follows with terms defined in units of angular acceleration:

$$\ddot{\underset{\sim}{\theta}} + M_A^{-1}B_A\dot{\underset{\sim}{\theta}} + \quad M_A^{-1}\underset{\sim}{\tau}_{As} \quad + M_A^{-1}S_A\dot{\underset{\sim}{\theta}}_p + M_A^{-1}\underset{\sim}{\tau}_{Aa} = \underset{\sim}{0} \quad (11)$$

or

$$M_A^{-1}K_A\Delta\underset{\sim}{\theta}$$

Just as we considered equation 8 to define a network of task-dynamical relations over a kinematic arm model, we also consider equation 11 to define an articulator dynamic network (articulator network) of relations among our (simplified) real arm's joint variables.

The task dynamic problem for our reaching example (and other real world tasks as well) may now be posed as the question of how to specify patterns of articulator dynamic controls (Equations 10 and 11) such that the resultant terminal device's free and perturbed kinematics evolve according to constraints embodied in the corresponding task space's topological dynamics (Equation 5). We consider two related methods in the sections below based on alternate versions of equation (11). The first method uses equations 8 and 11 to formulate task-specific equations of constraint or control laws over the articulator dynamic parameters; the second combines the use of control laws with the concept of network coupling between the task (equation 8) and articulator (equation 11) networks. Both methods address the issue of coordination in artificial (robotic, prosthetic) linkage systems. The network coupling method also affords a novel perspective on styles of control in physiological systems. In the following section, the control law approach is described, while the network coupling method will be discussed in a later section on physiological modes of motor control.

C. Method 1: Control laws

This method is conceptually quite simple and is outlined in Figure 6. First, one assumes that the model arm state $(\underset{\sim}{\phi},\dot{\underset{\sim}{\phi}})$ equals the real arm state $(\underset{\sim}{\theta},\dot{\underset{\sim}{\theta}})$ and that $\underset{\sim}{\theta}$ and $\dot{\underset{\sim}{\theta}}$ (hence, also, $\underset{\sim}{\phi}$ and $\dot{\underset{\sim}{\phi}}$) are specified proprioceptively. Second, one uses the following version of equation (11):

$$\ddot{\underset{\sim}{\theta}} + M_A^{-1}B_A\dot{\underset{\sim}{\theta}} + M_A^{-1}\underset{\sim}{\tau}_{As} + M_A^{-1}S_A\dot{\underset{\sim}{\theta}}_p + M_A^{-1}\underset{\sim}{\tau}_{Aa} = \underset{\sim}{0} \quad (12)$$

Third, by comparing equations 8 and 12, one can see that the real arm ($\underset{\sim}{\theta}$ variables) will move according to task dynamic requirements (i.e., will move identically to the task network's model arm [$\underset{\sim}{\phi}$ variables]) when the following identities hold: a) $J^{-1}M_B^{-1}B_BJ = M_A^{-1}B_A$; b) $J^{-1}M_B^{-1}K_B\Delta\underset{\sim}{x}(\underset{\sim}{\theta}) = M_A^{-1}\underset{\sim}{\tau}_{As}$; and c) $J^{-1}V\dot{\underset{\sim}{\phi}}_p = M_A^{-1}S_A\dot{\underset{\sim}{\theta}}_p + M_A^{-1}\underset{\sim}{\tau}_{Aa}$. Finally, one uses these identities to define the following nonlinear, state-dependent, articulator dynamic control laws:

$$B_A = M_AJ^{-1}M_B^{-1}B_BJ \quad (13a)$$

$$\underset{\sim}{\tau}_{As} = M_AJ^{-1}M_B^{-1}K_B\Delta\underset{\sim}{x}(\underset{\sim}{\theta}) \quad (13b)$$

$$\underset{\sim}{\tau}_{Aa} = (M_AJ^{-1}V-S_A)\dot{\underset{\sim}{\theta}}_p \quad (13c)$$

It should be noted that the articulator dynamic controls in equations 13 are defined by the linkage configuration($\underset{\sim}{\theta}$)- or state($\underset{\sim}{\theta}$,$\dot{\underset{\sim}{\theta}}$)- dependent products of: a) $M_A, S_A, J, J^{-1}, V, \underset{\sim}{x}(\underset{\sim}{\theta})$, and $\dot{\underset{\sim}{\theta}}_p$ --these are $\underset{\sim}{\theta}$- or $\dot{\underset{\sim}{\theta}}$ -dependent, but task independent;[7] and b) M_B^{-1}, B_B, K_B, and $\underset{\sim}{x}_o$---these are constant, but are dependent on both spatial context and task. Finally, one should note that for purposes of simplicity we have assumed that the computations involved in equations 13 occur instantaneously. However, in reality this cannot be the case and hence there must be a delay (Δt) between sensing a given linkage state (at $t = t_i$) and the specification of a task- and context-specific set of controls (at $t = t_i + \Delta t$). It is possible, therefore, that these controls will be totally inappropriate for the current ($t = t_i + \Delta t$) linkage state. There are two main ways to deal with this problem. The first is to minimize Δt by using a variety of methods: a) table lookup (e.g., Raibert, 1978) for those terms in equations 13 that are independent of the current spatial and task contexts, but can be indexed according to current articulatory state; b) parallel computation procedures, such that all elements in all matrices in (13) are not computed sequentially; c) computation strategies that heuristically omit certain terms in (13)[8] or that capitalize on the repeated use of certain "modular" functions (e.g., Benati, Gaglio, Morasso, Tagliasco, & Zaccaria, 1980) in the component terms in (13); and/or d) using remote sensing (exproprioception, e.g., vision) to specify certain kinds of information directly (e.g., hand position $\underset{\sim}{x}$) rather than indirectly through computations based on proprioceptive feedback (e.g., $\underset{\sim}{x}(\underset{\sim}{\theta})$). The second way of reducing the adverse consequences of delays is to use a predictive, "lookahead" type of computation (e.g., Ito, 1982; Pellionisz & Llinas, 1979) such that given an estimate of delay Δt, the system might sense a linkage state at $t=t_i$, predict the state at $t = t_i + \Delta t$, and perform equation 13's computations with reference to this predicted state.

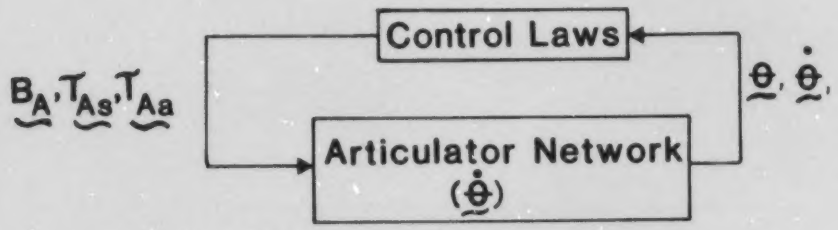


Figure 6. Overview of information flow in control law version of task dynamics.

V) Further Examples

In the preceding sections we described the details of the control law version of the task dynamic approach in the context of a discrete reaching

task's point attractor topology. In the present section, we generalize this approach to other task types as well as to variations on the discrete reach task theme. More specifically we describe how the task dynamic model: a) generates task specific trajectory shapes in discrete reaching, rhythmic target-to-target, cup-to-mouth, and crank-turning tasks; and b) provides "immediate compensation" to a sustained perturbation introduced to an effector system while en route to a target in a reaching task.

In the current control law context, all examples and computer simulations described below represent motions of the articulator network (the "real" arm), and the task network (the "model" arm) is rigidly constrained to move identically due to the assumptions that $\underset{\sim}{\phi} = \underset{\sim}{Q}$ and $\dot{\underset{\sim}{\phi}} = \dot{\underset{\sim}{Q}}$, given the current "proprioceptively" specified $\underset{\sim}{Q}$ and $\dot{\underset{\sim}{Q}}$.

A. Trajectory Shaping

1. Discrete reaching. This is the familiar reaching example, whose task space is defined as a two-dimensional point attractor (see Figure 4A). A straight-line trajectory for the terminal device (the hand) generated by these task dynamics for a discrete reach is illustrated in Figure 7 (trajectory a). For this trajectory the task space axis stiffnesses are symmetrical (i.e., $k_{T1} = k_{T2}$) and critical damping is assumed along both axes. Note, however, that perfect straight line trajectories are generated in contrast to the quasi-straight line trajectories observed experimentally for primates (e.g., Georgopoulos, Kalaska, & Massey, 1981; Morasso, 1981; Soechting & Lacquanti, 1981).

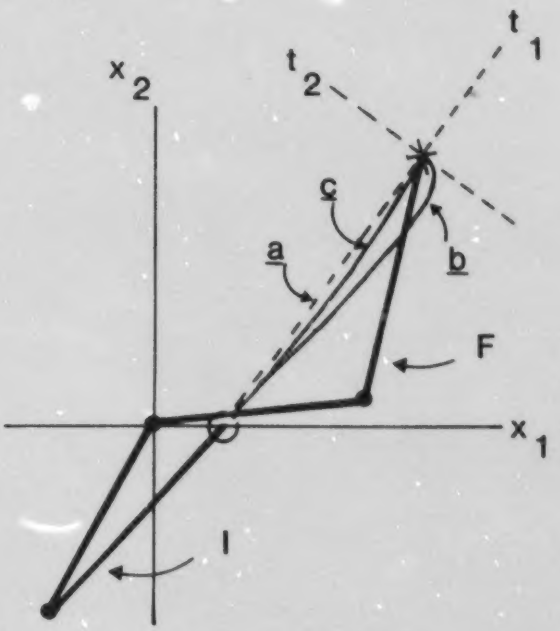


Figure 7. Body space discrete reaching trajectories showing effects of omitting velocity product torque compensation terms with different task axis weightings. I and F denote initial and final arm configurations, respectively.

As alluded to earlier (see also footnote 8), it is possible to omit $\mathcal{T}_{Aa}$ (i.e., the control vector associated with velocity product torques) from equation 12 and thereby obtain more "realistic" trajectories while at the same time reducing the amount of computation involved in specifying constraints on articulator dynamic parameters (trajectory b in Figure 7). As with trajectory a, trajectory b illustrates a reach involving symmetrical task axis stiffnesses and critical task space damping. Omitting $\mathcal{T}_{Aa}$ results in an articulator network whose velocity product terms are simply those specified by passive arm mechanics (i.e., $S_A\dot{Q}_D$ in equation 9) rather than those specified by the task network (i.e., $M_B\dot{V}_{q_D}$ in equation 7). Note that although the omission of $\mathcal{T}_{Aa}$ introduces a "hook" into trajectory b's illustrated hand motion, the hand nevertheless arrives precisely on target due to the underlying task space point attractor dynamics. This preservation of accurate targeting behavior when control terms related to velocity product torques are ignored is a feature of the task dynamic approach not shared by some other robotic control schemes (e.g., Hollerbach & Flash, 1981, see their Figure 8). Finally, it should be noted that straight line hand trajectories can be approximated when $\mathcal{T}_{Aa}$ is omitted by a judicious relative weighting of task axis stiffnesses. Hand trajectories progressively closer to ideal straight lines will be produced using progressively greater penalities for task mass deviations from the taskspace reach axis (t_1) en route to the target. A hand trajectory for the arm motion corresponding to one such ratio ($k_2:k_1=1.75:1$) with critical damping along both task axes is illustrated in Figure 7 (trajectory c).

2. Cup-to-mouth task. In a cup-to-mouth task the goal is to move a cup of liquid from an initial to final position (e.g., table top to mouth) while maintaining a horizontal spillage-preventing cup orientation during the movement. As in our discussion of the discrete reaching task, we begin with a simplified task-dynamic treatment of a planar cup-to-mouth task performed by a 3-joint (shoulder, elbow, wrist) arm using an abstract, functional description of that skill's task space. This task space is modeled as a three dimensional (one rotational and two linear degrees of freedom) point attractor and is illustrated in Figure 8A. In this figure the terminal device is an abstract task-segment (m_T = mass, l_T = length) representing the grasped cup, with one end (the "distal" end) defined as the point of final cup-mouth contact, and requiring three coordinates for its complete task space description. The target location (mouth) for the segment's distal end defines the origin (t_{01}, t_{02}) of a t_1t_2 Cartesian coordinate sysem; axis t_1 is defined as a reach-axis from the initial position of the segment's distal end to the t_1t_2 origin; and axis t_2 is defined orthogonally to t_1. The orientation of the task segment relative to axis t_1 defines the current angular t_3 coordinate; t_{03} defines the (identical) initial and target task segment orientations; and $I_T(=[1/3]m_Tl_T^2)$ is the task segment's moment of inertia about its distal end. The equations of motion corresponding to axes t_1, t_2, and t_3 are:

$$m_T\ddot{t}_1 + b_{T1}\dot{t}_1 + k_{T1}t_1 = 0 \qquad (14a)$$

$$m_T\ddot{t}_2 + b_{T2}\dot{t}_2 + k_{T2}t_2 = 0 \qquad (14b)$$

$$\rho I_T\ddot{t}_3 + \rho b_{T3}\dot{t}_3 + \rho k_{T3}(t_3-t_{03}) = 0 \qquad (14c)$$

where ρ is a constant scaling factor with units of length and is used to ensure dimensional homogeneity along all task space degrees of freedom. Thus,

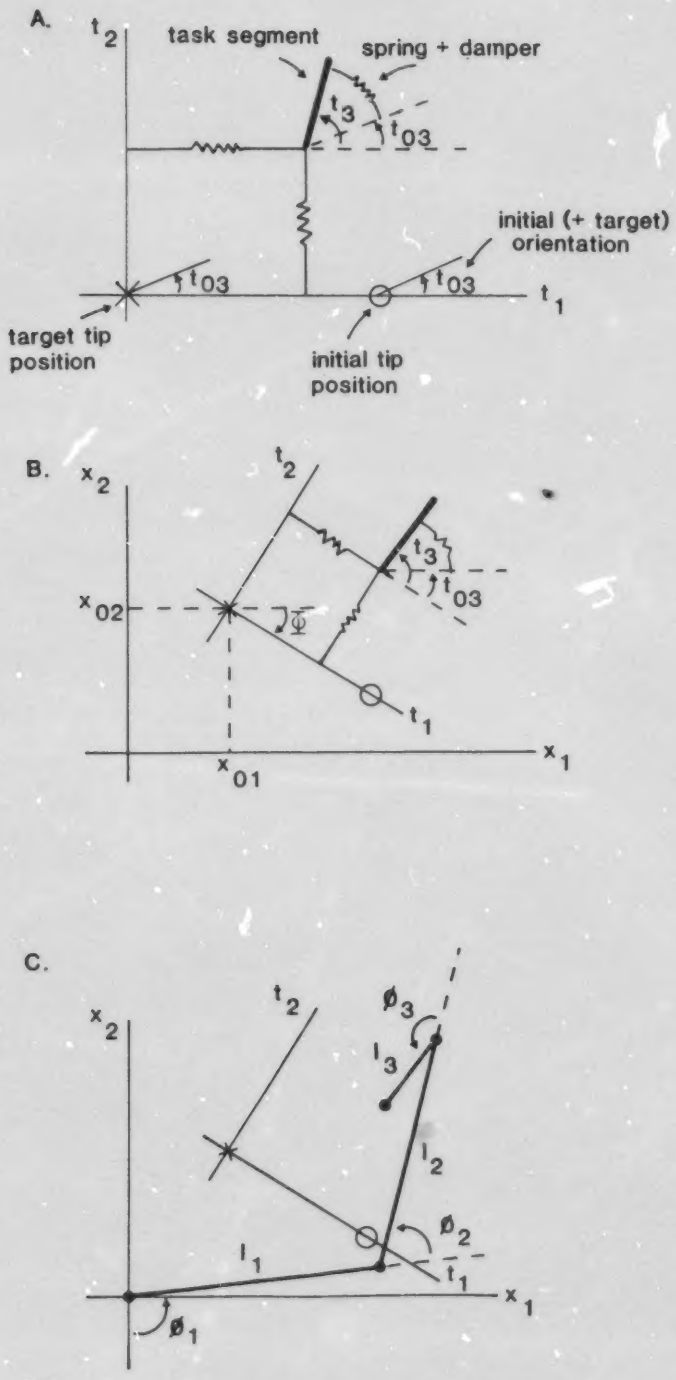


Figure 8. Cup-to-mouth task: A. Task space; B. Body space; C. Task network.

all terms of equation 14, even the rotational terms of 14c, are defined in units of force. For purposes of the present paper, ρ is set to 1 and consequently is omitted for notational simplicity in all further discussions in this section. In Figure 8A the stiffness and damping elements are represented in lumped form as squiggles in the lines connecting the task segment to the linear "rest positions" and to the rotational "rest orientation." Equations 14 describe a set of uncoupled (by definition of the abstract task space) equations with constant task dynamic parameters and can be represented in matrix form as:

$$M_T\ddot{\underset{\sim}{t}} + B_T\dot{\underset{\sim}{t}} + K_T\underset{\sim}{\Delta t} = \underset{\sim}{0}, \text{ where} \qquad (15)$$

M_T, B_T, and K_T are 3 x 3 diagonal matrices of task dynamic parameters analogous to the simpler 2 x 2 point attractor system of equation 5. In a similar fashion, the body spatial equation and the joint variable (task network) equation are simply the 3 x 3 analogs of equations 6 and 7. The corresponding body spatial and joint variable representations are illustrated in Figures 8B and 8C.

When simulated, a typical movement generated by these task dynamics, using symmetrical task axis stiffnessess ($k_{T1}=k_{T2}=k_{T3}$) and critical damping along all task axes, shows both a straight line trajectory and a maintained horizontal orientation of the task segment during the movement.

3. Reaching (rhythmic). The point attractor task space topologies used for the discrete reaching and cup-to-mouth tasks will be unable to generate the arm kinematics associated with sustained cyclic hand motion between two body spatial targets. Consider, for example, the case of planar motion of the terminal device (hand) and a corresponding 2-joint effector system (arm with shoulder and elbow joints). The task space is illustrated in Figure 9A and consists of an orthogonal pair of axes (t_1,t_2) for which: a) t_1 is defined along the line between the two targets (D=distance between the targets); and b) the origin is located midway between the two targets (A=D/2=distance from origin to either target). The terminal device is an abstract point task-mass (m_T=mass), and may be located anywhere in the task space. Point attractor dynamics are defined along axis t_2 to bring the task mass onto axis t_1 and to maintain it there despite transient perturbations introduced perpendicular to t_1. Limit cycle (periodic attractor) dynamics are defined along axis t_1 to sustain a cyclic motion of the task mass parallel to t_1 between the two targets, and to maintain the desired oscillation amplitude (A=D/2) despite perturbations introduced parallel to t_1. The task space equations of motion are:

$$m_T\ddot{t}_1 - b_{T1}\dot{t}_1 + c_{T1}t_1^2\dot{t}_1 + k_{T1}t_1 = 0 \qquad (16a)$$

$$m_T\ddot{t}_2 + b_{T2}\dot{t}_2 + k_{T2}t_2 = 0, \text{ where} \qquad (16b)$$

m_T, k_{T1}, b_{T2}, and k_{T1} are defined as in equation 5 (discrete reaching task, point attractor); and $(-b_{T1}\dot{t}_1 + c_{T1}t_1^2\dot{t}_1)$ is the nonlinear escapement term (van der Pol type) for axis t_1.

The dynamic parameters for axis t_2 are tuned in the same manner as in the t_2 axis of the discrete reach task space (see earlier Task space section). Tuning the dynamic parameters along axis t_1 involves specifying k_{T1} according to the desired period, P, of motion and the relation

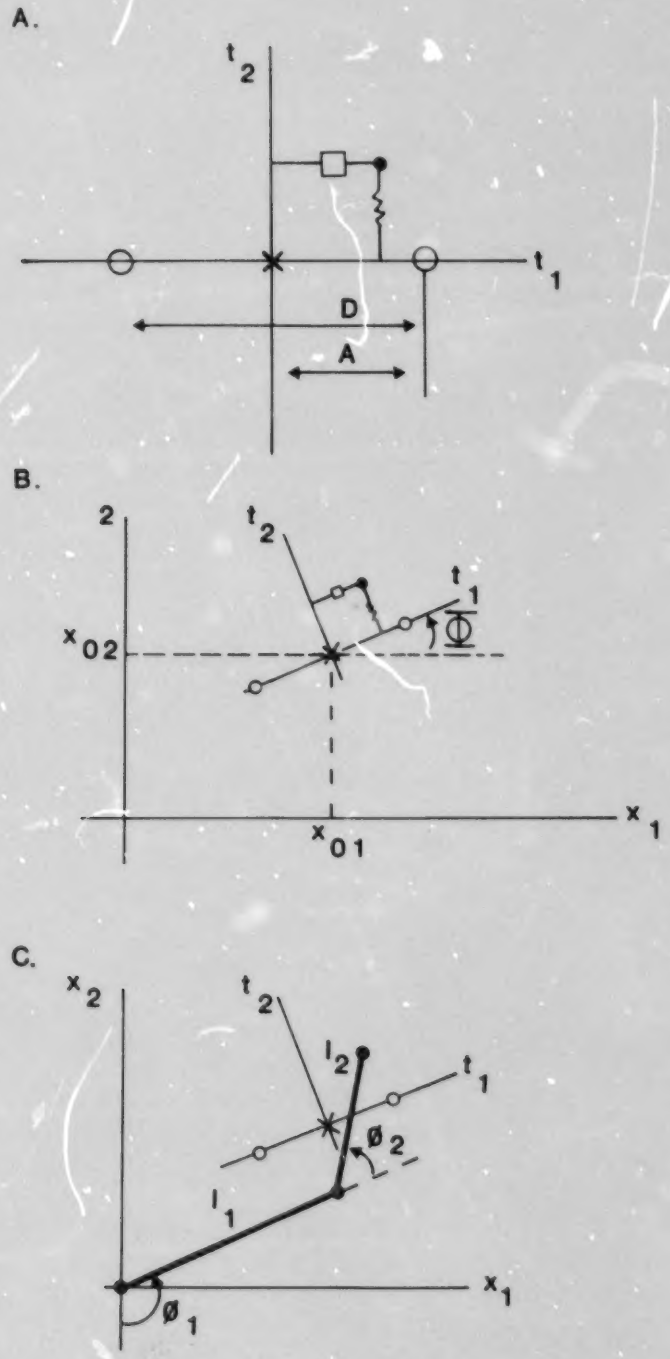


Figure 9. Rhythmic reaching: A. Task space. Open circles represent targets. Squiggle represents point attractor (spring and damper) dynamics along axis t_2. Open box represents limit cycle (spring and van der Pol escapement) dynamics along axis t_1; B. Body space; C. Task network.

$P=2\pi/\sqrt{k_{T1}/m_T}$. The procedure for specifying b_{T1} and c_{T1} is more involved, and may be understood by considering equation 16a in the following normalized, dimensionless form:

$$Z_1'' - \epsilon(1-Z_1^2)Z_1' + Z_1 = 0, \text{ where} \qquad (17)$$

a) the single and double apostrophe superscripts denote differentiation with respect to the dimensionless time variable, $N=\omega_0 n$, with $\omega_0=\sqrt{k_{T1}/m_T}$ and n denotes the standard time variable; b) $Z_1=\sqrt{c_{T1}/b_{T1}}\,t_1$ is the dimensionless displacement variable; and c) $\epsilon = b_{T1}/\sqrt{m_T k_{T1}}$ is a dimensionless measure directly related both to escapement "strength" (i.e., the strength with which the system resists being displaced from the limit cycle) and the shape of the limit cycle orbit in the phase plane (e.g., $\epsilon \ll 1$ corresponds to a circular orbit and sinusoidal motion; $\epsilon \gg 1$ corresponds to a relaxation orbit and a step-like motion). Given values for k_{T1} (from the desired period) and ϵ (from the desired orbit shape), b_{T1} is determined from the above expression for ϵ. Finally, it is known that the amplitude for the normalized (Z-variable) system of equation 17 is equal to 2.0 over a wide range of ϵ values (e.g., $0 \leq \epsilon \leq 10$, Jordan & Smith, 1977). For the original non-normalized (t-variable) system of equation 16a, the corresponding amplitude is $A = 2\sqrt{b_{T1}/c_{T1}}$. Therefore, given values for b_{T1} and desired amplitude A, the value of c_{T1} is determined from the preceding expression for A. Finally, equation 16 may be rewritten in matrix form as:

$$M_T\ddot{\underset{\sim}{t}} + B_T\dot{\underset{\sim}{t}} + K_T\underset{\sim}{t} = \underset{\sim}{F}_T, \text{ where} \qquad (18)$$

M_T and K_T are defined as in equation 5;

$B_T = \begin{bmatrix} -b_{T1} & 0 \\ 0 & b_{T2} \end{bmatrix}$, denoting the linear damping components; and

$\underset{\sim}{F}_T=(-c_{T1}t_1^2\dot{t}_1,\ 0)^T$, denoting nonlinear system components.

Equations 16 and 18 represent an autonomous, uncoupled, task spatial dynamical system with constant parameters. Figure 9B illustrates how the task space is located and oriented in body (shoulder) space. The body spatial dynamical system is described by:

$$M_B\ddot{\underset{\sim}{x}} + B_B\dot{\underset{\sim}{x}} + K_B\underset{\sim}{\Delta x} = \underset{\sim}{F}_B, \text{ where} \qquad (19)$$

$M_B=M_TR$, where R=the rotational transform matrix with elements r_{ij} defined previously in equation 6;

$B_B=B_TR$;

$K_B=K_TR$; and

$$\underset{\sim}{F}_B=(-c_{T1}(r_{11}\Delta x_1 + r_{12}\Delta x_2)^2\,(r_{11}\dot{x}_1 + r_{12}\dot{x}_2),\ 0)^T.$$

Equation 19 describes an autonomous, coupled (due to the rotation transformation), body spatial dynamical system with a constant set of linear parameters and a nonlinear, state-dependent forcing function. This body spa-

tial equation may be transformed kinematically into joint variable form by expressing $\underset{\sim}{x}$ variables as functions of the $\underset{\sim}{\phi}$ variables of a corresponding model arm (see Figure 9C):

$$M_B J\ddot{\underset{\sim}{\phi}} + B_B J\dot{\underset{\sim}{\phi}} + K_B \underset{\sim}{\Delta x}(\underset{\sim}{\phi}) = \underset{\sim}{F}_{\phi} - M_B V\dot{\underset{\sim}{\phi}}_p, \text{ where} \qquad (20)$$

M_B, B_B, and K_B are defined as in equation 19;

$J=J(\underset{\sim}{\phi})$, the Jacobian matrix;

$\underset{\sim}{\Delta x}(\underset{\sim}{\phi})=\underset{\sim}{x}(\underset{\sim}{\phi})-\underset{\sim}{x}_o$;

$\underset{\sim}{F}_{\phi} = \underset{\sim}{F}_B(\underset{\sim}{\phi})$, i.e., $\underset{\sim}{F}_B$ with the substitutions $\Delta x_1 = \Delta x_1(\underset{\sim}{\phi})$, $\Delta x_2 = \Delta x_2(\underset{\sim}{\phi})$, $\dot{x}_1 = J_{11}\dot{\phi}_1 + J_{12}\dot{\phi}_2$, $\dot{x}_2 = J_{21}\dot{\phi}_1 + J_{22}\dot{\phi}_2$; and

V and $\dot{\underset{\sim}{\phi}}_p$ are as defined in equation 7.

Equation 20 may be rewritten in the following task network form:

$$\ddot{\underset{\sim}{\phi}} + J^{-1}M_B^{-1}B_B J\dot{\underset{\sim}{\phi}} + J^{-1}M_B^{-1}K_B\underset{\sim}{\Delta x}(\underset{\sim}{\phi}) = J^{-1}M_B^{-1}\underset{\sim}{F}_{\phi} - J^{-1}V\dot{\underset{\sim}{\phi}}_p \qquad (21)$$

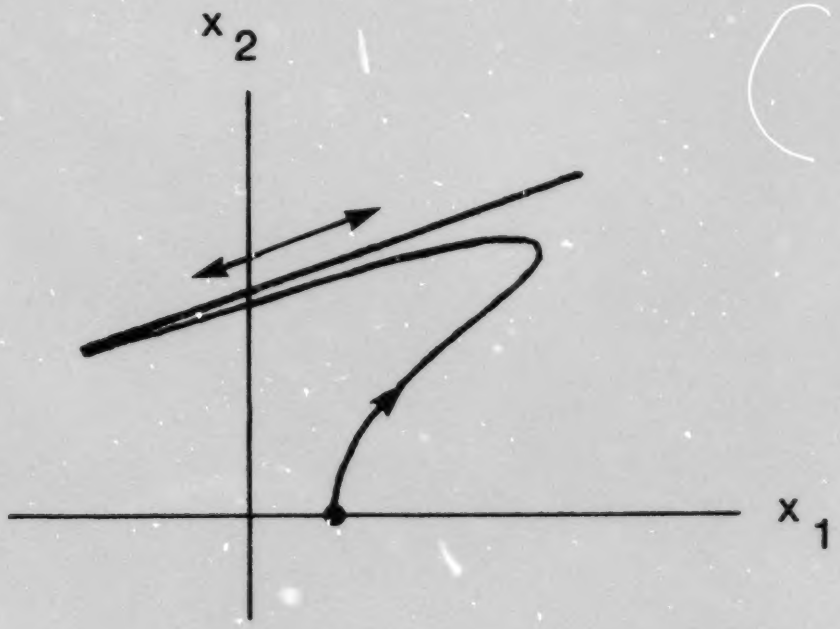


Figure 10. Body space rhythmic reaching trajectory when hand starts (or is perturbed to) a position away from the steady state trajectory.

Finally, since the real arm's motion can be described by the articulator network ($\underset{\sim}{\phi}$ variables) equation 12, one sees that the articulator controls, B_A

and τ_{As} are specified according to control laws 13a and 13b, respectively. A comparison of equations 12 and 21 shows that, assuming $\underset{\sim}{\phi}=\underset{\sim}{\theta}$ and $\dot{\underset{\sim}{\phi}}=\dot{\underset{\sim}{\theta}}$, the articulator control τ_{Aa} is defined according to the following control law:

$$\tau_{Aa} = (M_A J^{-1} \dot{V} - S_A)\dot{\underset{\sim}{\phi}}_p - M_A J^{-1} M_B^{-1} F_{\underset{\sim}{\phi}} \qquad (22)$$

A typical movement generated by these task-dynamics is illustrated in Figure 10, showing the motion of the task mass in body (shoulder) space. Note the straight line hand trajectory during the steady-state cyclic motion between targets, and also the way the hand is attracted autonomously to this steady-state trajectory despite a startup position (with zero velocity) away from this trajectory.

4. Crank Turning. Figure 11C illustrates the shoulder spatial layout of a crank turning task in which: a) motion of arm and crank occur in the horizontal plane; b) the crank segment's "distal" end is attached to a fixed rotation axis located at $\underset{\sim}{x}_o$ in shoulder space; c) the crank rotates at a constant angular velocity, V, about the fixed axis; d) the wrist joint is fixed and the hand tightly grasps the crank's handle, which freely rotates about an axis fixed to the crank's "proximal" end; and e) ϕ_1 and ϕ_2 represent the shoulder and elbow angles, respectively, while ϕ_3 represents the angle between the hand-forearm and crank. The task space description is illustrated in Figure 11A in which: a) the crank is the terminal device or task segment (m_T=mass, l_T=length); b) the fixed rotation axis at the crank's distal end defines the origin of a Cartesian t_1t_2 coordinate system; c) an angular t_3 coordinate is defined by the orientation of the crank relative to axis t_1; and d) $I_T=(1/3)m_T l_T^2$ is the crank's moment of inertia about its distal end. The task spatial equations of motion are defined as:

$$m_T\ddot{t}_1 + b_{T1}\dot{t}_1 + k_{T1}t_1 = 0 \qquad (23a)$$

$$m_T\ddot{t}_2 + b_{T2}\dot{t}_2 + k_{T2}t_2 = 0 \qquad (23b)$$

$$\rho I_T\ddot{t}_3 - \rho b_{T3}\dot{t}_3 + \rho c_{T3}\dot{t}_3^3 = 0, \text{ where} \qquad (23c)$$

ρ is the same scaling factor used in equation 14c, and will be omitted from further discussions in this section for notational simplicity.

Equations 23a and 23b define point attractors whose corresponding damping and stiffness factors are represented in lumped form in Figure 13a, and which serve to maintain the crank's distal end at the task space origin. Since in the real world the crank is fixed to this axis, these axes may be weighted rather loosely (i.e., they may be assigned low values for k_{T1} and k_{T2}). Equation 23c needs a bit more explanation as it contains a limit cycle's escapement term (Rayleigh type escapement: $-b_{T3}\dot{t}_3 + c_{T3}\dot{t}_3^3$) but no spring term. The behavior associated with equation 23c is best understood by examination of its corresponding phase portrait (Figure 12). Here it can be seen that there are three steady states represented by lines parallel to the t_3 axis. The lines defined by $\dot{t}_3 = \pm V = \pm\sqrt{b_{T3}/c_{T3}}$ are stable steady states, and the line $\dot{t}_3 = 0$ denotes an unstable steady state. In other words, given any nonzero startup velocity in either the upper or lower half plane, the system will reach the corresponding positive or negative steady state angular velocity, $\pm V$. If, however, the system begins at any angular

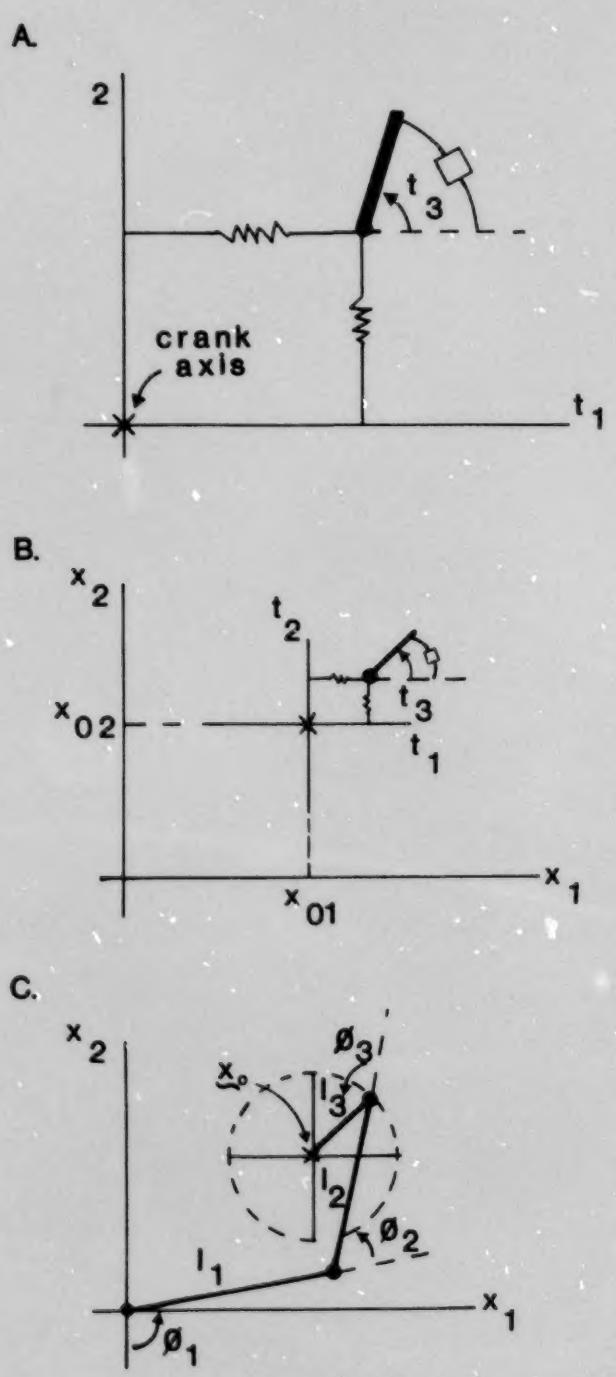


Figure 11. Crank turning: A. Task space. Squiggles represent point attractor dynamics along linear axes t_1 and t_2. Open box represents velocity attractor (Rayleigh escapement) dynamics along rotational axis t_3; B. Body space; C. Task network.

position with precisely zero velocity, it will simply stay at that position. In normalized form, equation 23c becomes:

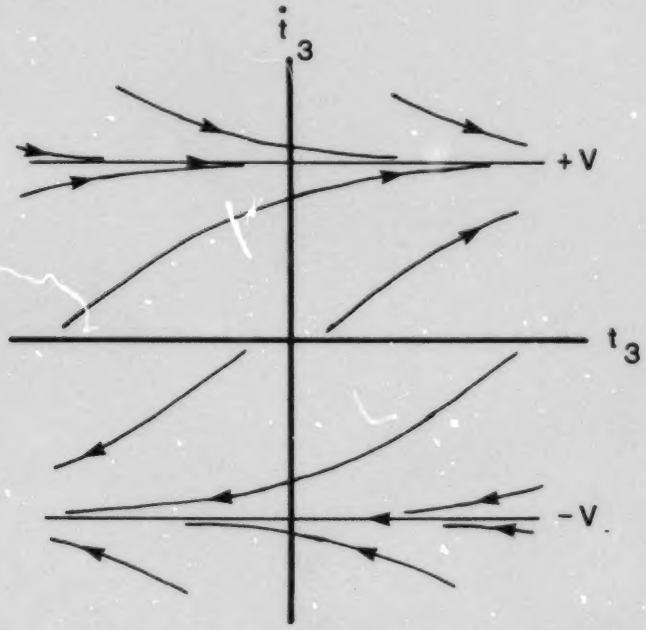


Figure 12. Phase portrait of velocity attractor system.

$$\ddot{Z}_3 - \epsilon(1-\dot{Z}_3^2)\dot{Z}_3 = 0, \qquad \text{where} \qquad (24)$$

$Z_3 = \sqrt{c_{T3}/b_{T3}}\, t_3$ is the dimensionless displacement variable, and $\epsilon = b_{t3}/I_T$ is directly related to the "strength" of the escapement (i.e., the speed with which the system attains the steady state and the strength with which it resists perturbations from the steady state). Since we are unaware of any other label for this type of dynamical topology, we will call it a bistable velocity attractor (or more simply, a velocity attractor). Given a desired escapement strength (ϵ) and final crank angular velocity (V), the above relationships are sufficient to tune the system's b_{T1} and c_{T1} values according to these task demands. Equations 23 may be rewritten in matrix form as:

$$M_T\ddot{\underset{\sim}{t}} + B_T\dot{\underset{\sim}{t}} + K_T\underset{\sim}{t} = \underset{\sim}{F}_T, \text{ where} \qquad (25)$$

M_T is defined as in equation 15;

$$B_T = \begin{bmatrix} b_{T1} & 0 & 0 \\ 0 & b_{T2} & 0 \\ 0 & 0 & -b_{T3} \end{bmatrix}; K_T = \begin{bmatrix} k_{T1} & 0 & 0 \\ 0 & k_{T2} & 0 \\ 0 & 0 & 0 \end{bmatrix}; \text{and}$$

$$\underset{\sim}{F}_T = (0,0,-c_{T3}\dot{t}_3^3)^T$$

Equations 23 and 25 represent an autonomous, uncoupled (by definition) task spatial dynamical system with constant parameters. Figure 11B shows how the task space is located and oriented in body (shoulder) spatial coordinates. Note that the orientation of task-to-body space is arbitrary, and in Figure 11B the orientation angle Φ is simply assumed to be zero (e.g., see Figure 8B for an example of a different task with nonzero Φ). Figure 11C, as mentioned previously, shows the relation of the task and body spaces to the task's "model" arm. Equations for body spatial, arm model, and task network dynamics may be derived from equation 25 in a manner similar to that used in generating equations 19, 20, and 21, respectively, from equation 18.

It should be noted that the configuration $\underset{\sim}{\phi}$ of the model arm is specified in exactly the same manner as our earlier examples. Angles ϕ_1 (shoulder) and ϕ_2 (elbow) can be obtained "propriceptively" but ϕ_3, the angle between the crank and the hand-forearm, cannot. However, assuming that the location of the crank's distal end (environmentally fixed rotation axis) is known in body space coordinates and given ϕ_1 and ϕ_2 proprioceptively, ϕ_3 is uniquely specified by geometric considerations. Thus the full $\underset{\sim}{\phi}$ set is available for use in the control law computations.

B. Immediate Compensation

In the Introduction, we reviewed experimental data on speech movements that showed task-specific, automatic, compensatory response patterns in remote articulators to unpredicted transient perturbations in a given articulator that were relatively immediate. These data implied that selective patterns of coupling or gating existed among the component articulators that were specific to the produced utterances. In the context of the task dynamic approach, we hypothesize that these coupling patterns are due to the corresponding evolving patterns of articulator-dynamic control parameters specified by task- and state-dependent control laws or equations of constraint.

To illustrate, consider the following example of a discrete reaching task (formulated as a modified version of a cup-to-mouth task) in which: a) the terminal device is a pointer fixed to the hand of a 3-segment (upper arm, forearm, hand-pointer) arm; b) planar motion of the pointer corresponds to angular motions of the arm's 3 joints (ϕ_1=shoulder, ϕ_2=elbow, ϕ_3-angle between pointer and forearm); and c) task demands focus on positioning the pointer's distal end at a body-spatial x_1x_2 target but are relatively indifferent to the precision of final orientation control. Consequently, the task space may be described as a 3-dimensional point attractor with symmetrical weightings for the linear t_1 and t_2 axes, and a much smaller weighting for the rotational t_3 axis. Figure 13 illustrates the initial (a) and final (b) arm configurations that correspond to the current task dynamics (weighting ratio of axes t_1 and t_2 to t_3 is 20:1) when the arm encounters no perturbations en route to its body spatial pointer target. The initial arm configuration is $\underset{\sim}{\theta}_i = (79^\circ, 20^\circ, 171^\circ)^T$ and the final arm configuration $\underset{\sim}{\theta}_f = (115^\circ, 81^\circ, 75^\circ)^T$. Figure 13 (configuration c) shows the final arm position when the shoulder angle is suddenly braked during the trajectory when it reaches 105° and is held fixed at this angle. The initial $\underset{\sim}{\theta}_i$ is the same as in the unperturbed case and the pointer's distal end reaches precisely the same spatial x_1x_2 target as in the unperturbed motion, despite the fact that the final configuration has changed to $\underset{\sim}{\theta}_f = (105^\circ, 95^\circ, 52^\circ)^T$. In other words, the system's response to the perturbation was to "automatically" redistribute the activity among its component degrees of freedom in a

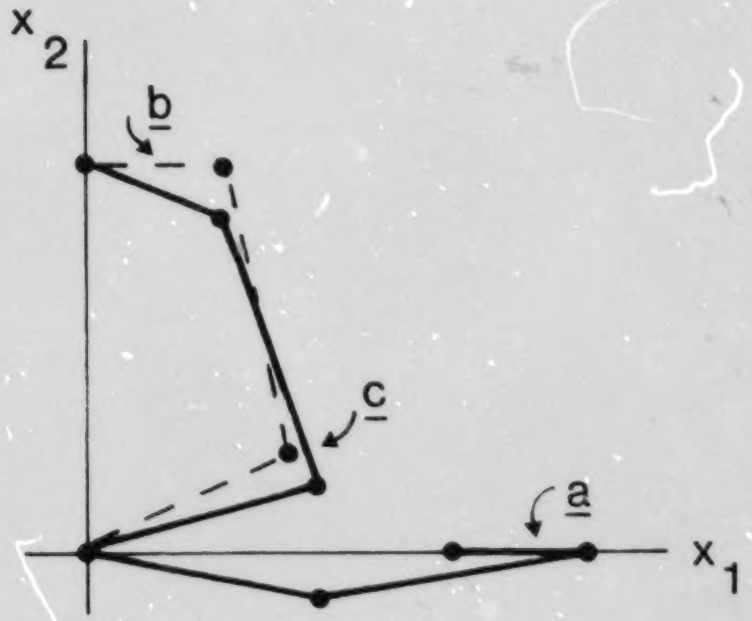


Figure 13. Arm configurations for simulated discrete reaches showing: a. Initial posture; b. Final posture (unperturbed trajectory); c. Final posture (perturbed trajectory).

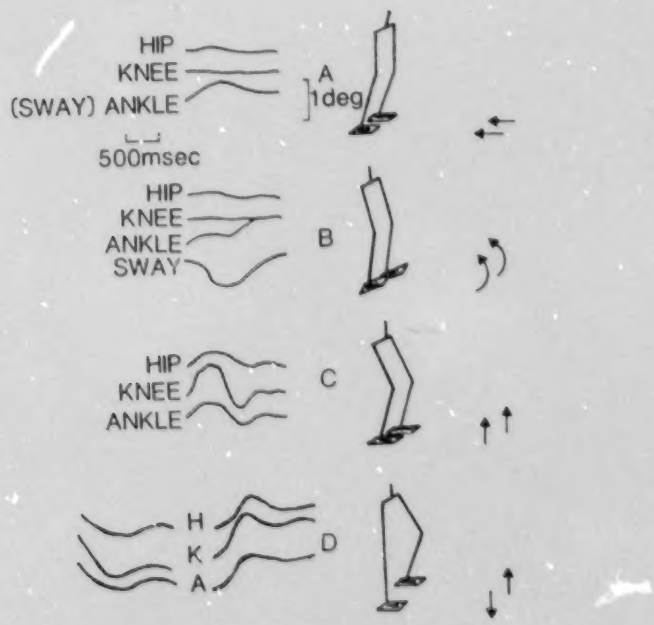


Figure 14. Description of basic postural perturbation paradigm of Nashner and colleagues, showing four types of perturbation (right column) and corresponding leg joint angular rotations (left column); A. AP translation; B. Direct rotation; C. Synchronous vertical; D. Reciprocal vertical. (Adapted from Nashner & Woolacott, 1979.)

manner that still achieved the same task spatial goal. Furthermore, such compensatory motor equivalence reflects the fact that targets in the task dynamic approach are not specified as final articulator configurations, but rather as desired final spatial coordinates for the terminal device; the final articulator configuration "falls out" of the task dynamic organization and the environmental conditions in which the movement is performed.

VI) Relevance to Physiological Literature

In this section, we will describe how the task dynamic approach might apply to the issue of postural control in humans, and suggest a "systemic" alternative to the modular synergy model of Nashner (e.g., Nashner, 1981; Nashner & Woolacott, 1979) to account for postural compensatory phenomena. Further, we will review evidence from studies of single-joint discrete movement tasks (e.g., Bizzi, Chapple, & Hogan, 1982) that show that the physiologically relevant parameter of rest angle (i.e., the angle specified by the equilibrium point between agonist and antagonist length-tension curves) is actively specified during such tasks as a gradually (as opposed to step-like) changing central control signal. In the control law version of task dynamics, there is no articulator-dynamic control parameter corresponding to rest angle. However, a network coupling version of task dynamics, now in preliminary form, will be described that includes rest angle as a parameter and provides a rational account for the evolution of the rest angle's trajectory without requiring an explicitly preplanned trajectory representation to account for the observed pattern. Finally we will discuss the implications of the network coupling approach for theories of learned complex skilled actions.

A. Postural control

Nashner and his colleagues have performed an elegant series of experiments on postural responses to support surface perturbations in standing human subjects. Summarizing from the experimental report of Nashner, Woolacott, and Tuma (1979) and several subsequent reviews (Nashner, 1979; Nashner, 1981; Nashner & Woolacott, 1979), we can describe the paradigm and findings in the following way. Basically, a subject stands with each foot on a separate horizontal platform that can be translated horizontally, translated vertically, or rotated about an axis aligned with the ankle joint. Using these platforms, one or a combination of the following four types of perturbation could be delivered to the subjects on a given trial (Figure 14): a) simultaneous forward or backward anteroposterior translation (AP translation); b) simultaneous flexion or extension rotations (direct rotation), c) simultaneous upward or downward vertical translation (synchronous vertical); and d) reciprocal vertical translation (reciprocal vertical). These perturbation types may be characterized by the corresponding patterns of whole body motions and joint rotations that would be induced in "passive" noncompensating subjects (Figure 14). Thus, AP translation caused the body to lean in the direction opposite to the translation; direct rotation caused the body to tilt in the same sense as the rotation; synchronous vertical caused the body to move with the translation; and reciprocal vertical caused the body to tilt laterally toward the lowering platform. It should be noted that the first three perturbation types induce motions in the sagittal plane, while the reciprocal vertical type induces motion in the frontal plane.

In response to each perturbation type or type combination, Nashner et al. measured EMG responses from the upper and lower leg muscles, as well as changes in ankle, knee, and hip angles. Associated with each perturbation type was a long latency (e.g., 100-110 ms latency in gastronemius) "rapid postural adjustment" (Nashner, 1981), which comprised the earliest useful postural response, while the shorter latency myotatic reflexes were either absent or of no apparent functional value. These rapid postural adjustments for a given type a) were characterized by fixed ratios of activity among the responding muscles, b) were specific to the perturbation type (and the corresponding type-specific patterns of joint displacements), and c) were "functionally related to the task of coordinating one kind of postural adjustment" (Nashner, 1979, p. 179). Further, during a set of trials in which a sequence of either of three perturbation types (AP translation, synchronous vertical, or reciprocal vertical) was unexpectedly and immediately followed by a sequence of one of the other two types, it was found that the functionally appropriate postural synergy response occurred even on the first trial of the new type. Such "first trial adaptation" did not occur, however, when a sequence of AP translations was followed immediately by an unexpected series of direct rotations (or vice versa). In these cases, the functionally correct (i.e., posturally stabilizing) synergistic response pattern was implemented progressively over a series of approximately three to five trials. Additionally (Horak & Nashner, 1983), if a series of AP trials with the subject standing directly on the footplates was followed by a series of AP trials with the feet resting on narrow transverse beams, the subjects switched from a postural response involving predominantly ankle motions (ankle strategy) to one involving predominantly hip motions (hip strategy). This strategy change was implemented progressively over the course of approximately 5-20 trials, and this multitrial adaptation process was also seen for the reverse change from beam to footplate postural strategies.

Nashner and his colleagues have interpreted these data as being consistent with a modular synergy "conceptual model for the organization of postural adjustments" (e.g., Nashner, 1979, 1981; Nashner & Woolacott, 1979). Although admittedly in preliminary form, this hierarchical model proposes that postural synergies are organized spinally as separate modular function generators, and are automatically triggered by correspondingly appropriate distinctive features of somatosensory (i.e., proprioceptive information related to joint angular rotations) inputs. Thus, for example, the AP sway synergy module is activated in proportion to ankle rotational input, while the vertical suspensory synergy module is activated in proportion to knee rotational input, and inhibition of the sway module by the suspensory module is provided to prevent simultaneous activation of both synergies. Such a system provides a reasonable account of the automatic first trial postural responses described above. Additionally, supraspinal processes are assumed to modulate the input-output relationships of the peripheral synergy modules in order to maintain postural stability using posturally relevant knowledge of results (e.g., sensory conflict between somatosensory and vestibular sources of information concerning the body's orientation relative to the support base and the line of gravity). Such supraspinally controlled modulation effects are presumed to occur relatively slowly, and are posited to underly the multitrial postural adaptation phenomena described above.

Task dynamics offers an attractive alternative to this hierarchical modular synergy approach. In the modular approach, synergies are canonically represented as stored output patterns, and are triggered by corresponding dis-

tinctive features of somatosensory inputs. When the problem of postural control is formulated in task dynamic terms, however, synergies need not be canonically represented anywhere; rather, synergistic patterns of muscle activity may be viewed as emergent properties of the task dynamically organized postural system. In this latter view, one may define a postural task space (see Figure 15A) in the following way, using postural control only in the sagittal plane for purposes of illustrative simplicity. This task space is modeled as a two-dimensional point attractor for which: a) the terminal device is the body's center of mass and is represented as a point mass with mass m_T equal to total body mass (note that, unlike earlier examples, this terminal device cannot in general be associated with a particular point on the linkage); b) axis t_2 is defined parallel to the line of gravity and axis t_1 is defined normal to t_2; the t_1t_2 origin is defined by the target location of the mass center, which coincides with the mass center's initial location (assuming a corresponding posturally stable initial body configuration). The task space equations of motions are:

$$m_T\ddot{t}_1 + b_{T1}\dot{t}_1 + k_{T1}t_1 = 0 \qquad (26a)$$

$$m_T\ddot{t}_2 + b_{T2}\dot{t}_2 + k_{T2}t_2 = 0, \text{ where} \qquad (26b)$$

the damping and stiffness parameters define point attractor topologies along each task axis. Gravity does not appear explicitly in (26b) since t_2 denotes displacement from the statically stable vertical position of the task mass in the gravitational field. In other words, $t_2 = t_2^* - (m_Tg/k_{T2})$, where t_2^* corresponds to the statically stable "vertical" position of the task mass in the absence of gravity, and g denotes the acceleration due to gravity. In matrix form these equations become:

$$M_T\ddot{\underline{t}} + B_T\dot{\underline{t}} + K_T\underline{t} = 0 \qquad (27)$$

The pattern of task spatial dynamic parameters in (27) may be transformed into body spatial form with reference to a coordinate system whose origin coincides with the center of the support base. The spatial relationships between task space and body space are illustrated in Figure 15B in which: a) the x_1 axis is defined along the anteroposterior line between the rear and front edges (denoted by open squares) of the support base, which is defined by the contact areas between the feet and ground surface; b) the x_2 axis is defined normal to x_1 at the midpoint of the support base; c) the relative orientation between task and body space is defined by the angle Φ; and d) the location of the task space origin in body space coordinates is defined by $\underline{x}_o$. It should be noted that both Φ and $\underline{x}_o$ are defined by the current postural configuration, which is assumed to be statically stable, i.e., the projection of the initial location of the center of mass (task space origin) along the line of gravity will fall within the boundaries of the support base. In this regard, the task dynamic approach to vertical posture control is similar to the model proposed by Litvintsev (1972), who stated that it is likely that "the essential role in equilibrium maintenance is played by a mechanism which organizes muscular control at the various joints by parameters characterizing the general body position...(p. 590)," and that "the magnitude and the rate of deviation of the weight center projection on the support plane are input parameters for this mechanism (p. 598)." Finally, it should be noted that: a) in most daily activities we stand on horizontal surfaces and Φ thereby usually assumes a value of zero (Figure 15C); and b) the body spatial

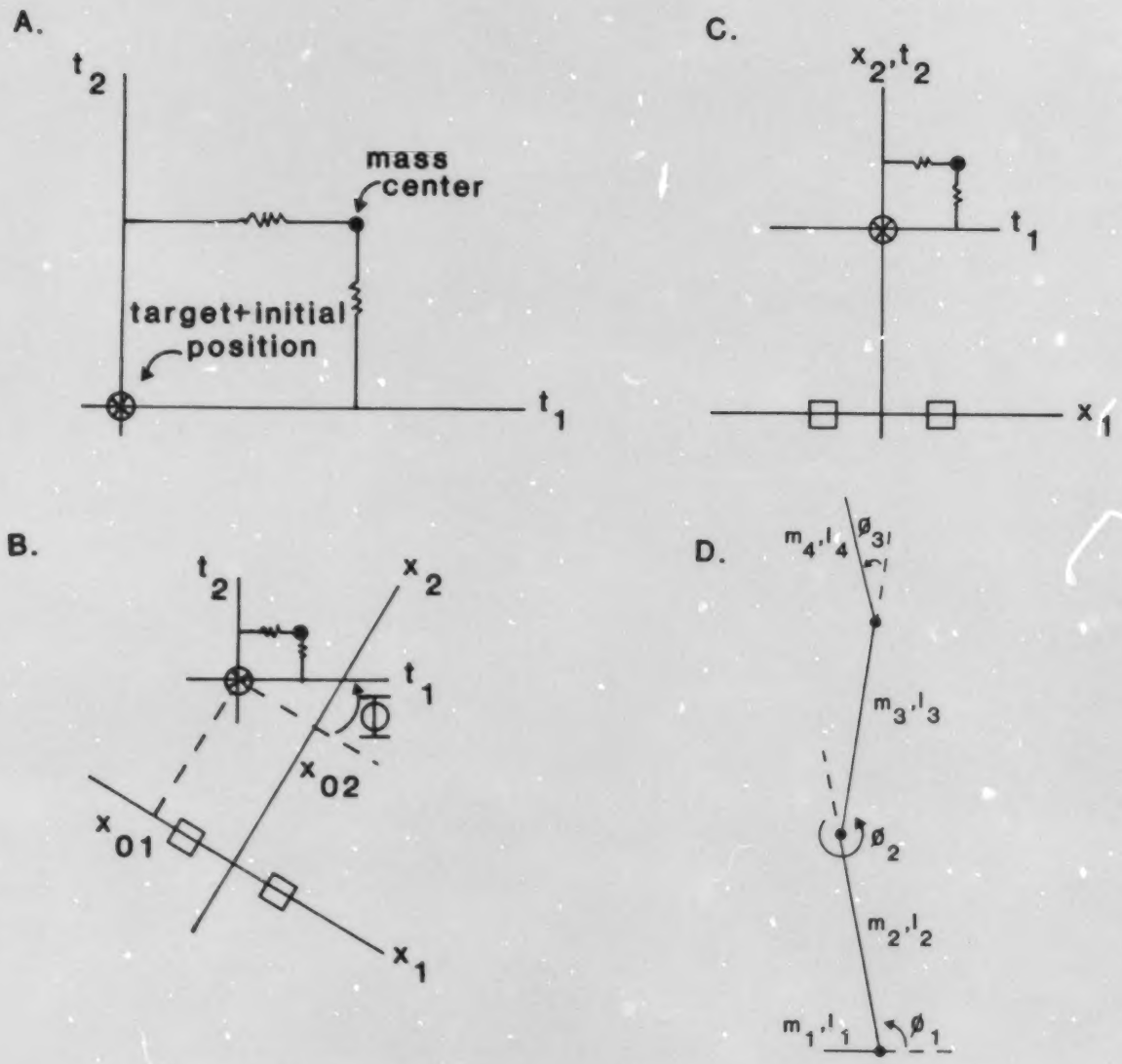


Figure 15. Postural maintenance task: A. Task space; B. Body space. Open boxes represent front and back edges of support base (feet). Orientation angle, Φ, between task and body space is nonzero; C. Body space. Φ is zero, representing parallel orientation of t_1 and x_1; D. Postural effector system.

equations of motion derived from (27) have the same form as equation (6) in our earlier discrete reaching example.

The body spatial pattern of dynamic parameters may be transformed into an equivalent task network expression based on the joint variables of a (simplified) four-segment (foot, shank, thigh, torso), three-joint (ankle, knee, hip) effector system (Figure 15D). This task network equation has the same general form as the discrete reaching equation (8), except that the postural task network involves three joints (not two joints), and two spatial variables, defining thereby a redundant task-articulator situation (see footnote 7) and hence requiring the use of the Jacobian pseudoinverse, J^+, or weighted pseudoinverse, J^*.

In the task dynamic framework, it is evident that consistent synergistic patterns of postural responses will occur in response to given types of destabilizing inputs. If a task network is established according to an accurate evaluation of the spatial relationships between task and body space, these postural responses will be stabilizing and compensatory. Further, they will be "immediately" accurate since they depend only upon the current limb state and the (accurately tuned) task network. In other words, synergistic responses emerge from the (tuned) postural system's underlying task dynamic organization; there is no need to invoke the notion of access to and triggering of stored canonical synergy output programs. However, the postural system can be fooled into establishing an improperly tuned task network based either on an inappropriate evaluation of the task-body space geometric relationship, or on the use of an inappropriate weighting strategy for the joints in the (redundant) postural effector system. In the former case, for example, a series of trials involving AP translation perturbations requires tuning $\mathbb{O} = 0$, since the support base is horizontal throughout the trials. If a direct rotation perturbation is unexpectedly introduced, this setting is no longer valid and the task network will shape postural responses that are inappropriate and destabilizing for the new task-body space geometry. Adaptive responses to direct rotation perturbations require setting $\mathbb{O} = \phi_1$ (where ϕ_1 = ankle angle) in order to tune the task network appropriately. Apparently, this sort of retuning process does not occur instantaneously, but requires 3-5 trials as discussed earlier.

In the case of tunings related to effector system weighting strategies, it appears that an efficient strategy for dealing with AP translation perturbations of the foot plates is an ankle-predominant one when the feet rest directly on the plates, but a hip-predominant one when the feet rest on narrow beams. These strategies would serve to tune differentially the weighting matrices for the task network (via the weighted pseudoinverse, J^*) according to the current support surface configuration. If the support surface context is changed, say, from plate to beam support, then the ankle-weighted J^* used for the plate context will be inappropriate for (or less efficient than) the new beam context. Apparently, adaptively retuning J^* to reflect a hip predominant strategy (and vice versa for hip to ankle strategy retuning) requires approximately 5-20 trials as discussed earlier.

B. Rest angle trajectories; Network coupling

1. Rest angles: final position control, trajectory formation. It was noted above (in the Topology and Dynamics section), that discrete target acquisition tasks in one degree of freedom systems (e.g., at the elbow joint)

were observed to display properties homologous to damped mass-spring systems by several investigators (e.g., Cooke, 1980; Fel'dman, 1966; Kelso, 1977; Polit & Bizzi, 1978; Schmidt & McGown, 1980) and had been modeled, essentially, as point attractors in an articulator dynamic sense, requiring only the setting of the final or target rest angle parameter (but see footnote 4). According to the so-called "final position control" hypothesis (e.g., Bizzi, Accornero, Chapple, & Hogan, 1981; Kelso & Holt, 1980; Sakitt, 1980), the relative levels of neural activation of the spring-like agonist and antagonist muscle groups at a joint

> define an equilibrium point between two opposing length-tension curves and consequently a joint angle. It has been suggested that the transition from a given position to another may occur whenever the CNS (central nervous system) generates a signal shifting the equilibrium point between the two muscles by selecting a new pair of length-tension curves (Bizzi et al., 1981).
>
> According to this schema, movements are, at the simplest level, transitions in posture. This simple idea is attractive because the details of the movement trajectory will be determined by the inertial and visco-elastic properties of muscles and ligaments around the joint (ibid).

However, as we discussed above (Section III), such an articulator-dynamic control scheme breaks down when more complex multijoint tasks are considered (see also footnote 4). Further, even for single degree of freedom positioning tasks, the final position control hypothesis may be incomplete. Bizzi and his colleagues (Bizzi & Abend, 1982; Bizzi et al., 1981; Bizzi, Accornero, Chapple, & Hogan, 1982; Bizzi, Chapple, & Hogan, 1982), for example, have suggested that the rest angle trajectory is controlled in addition to final position. Thus, the final position control hypothesis predicts that elbow movements result from rapid shifts to target equilibrium points and that, consequently, steady state equilibrium positions would be achieved after a delay from muscle activity onset due solely to the dynamics of muscle activation. Bizzi, Chapple, and Hogan (1982) offer a "slowest case" approximation of 150 ms for the time taken by the net muscle force to rise within a few percent of its final value. In fact, however, these investigators showed that for movements of at least 600 ms in duration, the mechanical expression of alpha motoneuronal activity reached steady state only after at least 400 ms had passed following the onset of muscle activity. Consequently, it appears that the centrally generated rest angle signal gradually changes during the movement, even in deafferented monkeys, such that the alpha motoneuronal activity defines "a series of equilibrium positions, which constitute a trajectory whose end point is the desired final position" (ibid). Finally, it should be noted that Bizzi et al. (1981) interpret their observations as implying the existence of trajectory plans or programs to account for the observed time courses of rest angle movement as well as the final rest angle position.

The control law version of task dynamics is unable to account for these data for two reasons. First, there is no parameter corresponding to rest angle in the single degree of freedom case or rest configuration in the multi-degree of freedom case. Second, the control law version assumes that $\underset{\sim}{\theta}$ and $\dot{\underset{\sim}{\theta}}$ (real arm state) are perceived proprioceptively, that $\underset{\sim}{\phi}$ and $\dot{\underset{\sim}{\phi}}$ (model arm state) equal the real arm's state, and that control laws are specified according to the currently perceived real arm's state. In the "deafferented" case,

in which the current $\underset{\sim}{\theta}$ and $\dot{\underset{\sim}{\theta}}$ are unavailable, the control laws are undefined and (coordinated) motion is not possible. Given the above "trajectory formation" data of Bizzi and colleagues, if task dynamics is to be applied in these situations, the control law version must be amended to generate coordinated movements in deafferented preparations and to include a rest configuration parameter (which, of course, must evolve autonomously during the movement according to task-dynamic constraints). Although in preliminary form, we believe a network coupling version of task dynamics satisfies these requirements and provides a more biologically plausible task dynamic account of skilled movements.

2. Network Coupling. The network coupling method (outlined in Figure 16) involves shaping articulator dynamics according to task-specific dynamical constraints and may closely approximate a biological style of coordination and regulation. Briefly, the network coupling method involves interpreting the observed skilled motion of an effector system to be the observable "output" of an articulator network that comprises, however, only one half of a task specific action system. The complete action system consists of the mutually or bidirectionally coupled task (output variables: $\underset{\sim}{\phi}$, $\dot{\underset{\sim}{\phi}}$, $\ddot{\underset{\sim}{\phi}}$, etc.) and articulator (output variables: $\underset{\sim}{\theta}$, $\dot{\underset{\sim}{\theta}}$, $\ddot{\underset{\sim}{\theta}}$, etc.) networks. Thus, for the multidegree of freedom discrete reaching task described earlier, this method involves: a) treating the task network defined in equation 8 as a system for intrinsic pattern generation that is specified for a given task and actor-environment context, and that does not require peripheral input for its operation; b) defining the articulator network corresponding to an actual arm by the following version of equation (11):

$$\ddot{\underset{\sim}{\theta}} + M_A^{-1} S_A \dot{\underset{\sim}{\theta}}_p + M_A^{-1} B_A \dot{\underset{\sim}{\theta}} + M_A^{-1} K_A \Delta\underset{\sim}{\theta} + M_A^{-1} \underset{\sim}{T}_{A_a} = \underset{\sim}{0} \quad (28)$$

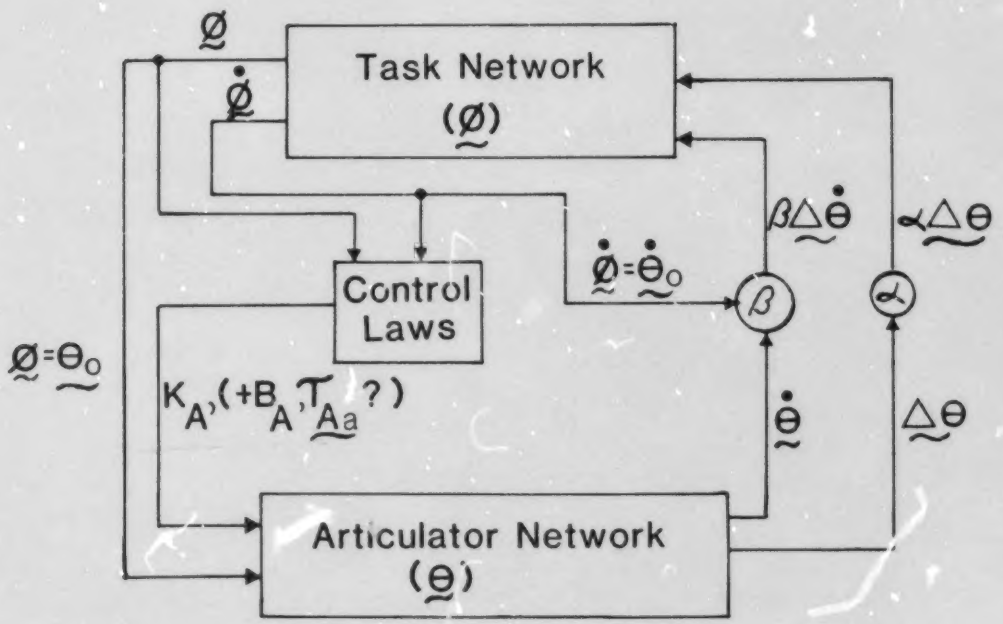


Figure 16. Overview of information flow in network coupling version of task dynamics.

and c) using the task network to both actuate and modulate the articulator network, while using the articulator network to modulate the task network.

More specifically, the network coupling method begins by using the $\underset{\sim}{\phi}$ output of the task network as the $\underset{\sim}{\Theta}_o$ input ("rest configuration," i.e., $\underset{\sim}{\Theta}_o = \underset{\sim}{\phi}$) for the articulator network. However, since the task and articulator networks are potentially independent, one cannot simply assume identical task and arm network states (as in the control law approach). Rather, we make the less stringent assumption that real arm and model arm states are "close," i.e., that $\underset{\sim}{\Theta} - \underset{\sim}{\phi} = \underset{\sim}{\Delta\Theta}$ and $\dot{\underset{\sim}{\Theta}} - \dot{\underset{\sim}{\phi}} = \dot{\underset{\sim}{\Delta\Theta}}$ are "small." Therefore, the constraint relationships for B_a, K_A, and $\mathcal{T}_{Aa}$ are defined in a more approximate sense than those in equation (13) (see Appendix C for details):

$$B_A = [M_A J^{-1} M_B^{-1} B_B J] \big|_{\underset{\sim}{\Theta}_o = \underset{\sim}{\phi}} \qquad (29a)$$

$$K_A = [M_A J^{-1} M_B^{-1} K_B J] \big|_{\underset{\sim}{\Theta}_o = \underset{\sim}{\phi}} \qquad (29b)$$

$$\mathcal{T}_{Aa} = [M_A J^{-1} V - S_A] \big|_{\underset{\sim}{\Theta}_o = \underset{\sim}{\phi}, \dot{\underset{\sim}{\phi}} = \dot{\underset{\sim}{\phi}}_p} \qquad (29c)$$

These sets of "driving" constraints ($\underset{\sim}{\Theta}_o = \underset{\sim}{\phi}$) and "modulating" constraints (equations 29) comprise the "efferent" aspect of our coupled-network action system. With these constraints, the articulator network becomes (statically) stable about the current rest configuration, with stiffness and damping properties defined relative to task space axis directions.[10]

However, a coupled-network action system involves bi-directional coupling and hence an "afferent" aspect as well. This pattern of afferentation serves to modulate the activity of the task network on the basis of both relative angular displacement ($\underset{\sim}{\Delta\phi} = \underset{\sim}{\phi} - \underset{\sim}{\Theta}$) and relative angular velocity ($\dot{\underset{\sim}{\Delta\phi}} = \dot{\underset{\sim}{\phi}} - \dot{\underset{\sim}{\Theta}}$) coupling terms defined by $\alpha \underset{\sim}{\Delta\phi}$ and $\beta \dot{\underset{\sim}{\Delta\phi}}$, respectively, where α and β are constant scalar coupling coefficients. This type of coupling, which is proportional to differences between corresponding sets of state variables, is called diffusive coupling (e.g, Rand & Holmes, 1980). The modulated task network is then described by the following amended version of equation (8):

$$\ddot{\underset{\sim}{\phi}} + J^{-1} M_B^{-1} B_B J \dot{\underset{\sim}{\phi}} + J^{-1} M_B^{-1} K_B \underset{\sim}{\Delta x}(\underset{\sim}{\phi}) + J^{-1} V \dot{\underset{\sim}{\phi}}_p + \alpha \underset{\sim}{\Delta\phi} + \beta \dot{\underset{\sim}{\Delta\phi}} = \underset{\sim}{0}, \quad (30)$$

where by assumption $\underset{\sim}{\Delta\phi}$ and $\dot{\underset{\sim}{\Delta\phi}}$ are assumed "small." The effects of these coupling terms on system behavior are to reduce the size of $\underset{\sim}{\Delta\phi}(= -\underset{\sim}{\Delta\Theta})$ via $\alpha \underset{\sim}{\Delta\phi}$ coupling and to reduce the size of $\dot{\underset{\sim}{\Delta\phi}}$ via $\beta \dot{\underset{\sim}{\Delta\phi}}$ coupling, thereby promoting an in-phase (vs. anti-phase) one-to-one relationship between real and model arm motions. It should be noted that equation (30) reverts to equation (8) when $\underset{\sim}{\Delta\phi}$ and $\dot{\underset{\sim}{\Delta\phi}}$ equal zero (i.e., there is perfect mutual tracking of the real and model arms) or when the afferent coupling is disengaged (i.e., peripheral feedback is eliminated and the system is "deafferented") by setting α and β to zero. Further, one should note that, even when deafferented, the model arm is governed by the task network equation (30) and hence $\underset{\sim}{\Theta}$ is capable of coordinated (although probably degraded) motion due to "internal feedback" of the model arm's current state within the task network. Here, internal feedback is used in the sense of Evarts (1971) to indicate information "arising from structures within the nervous system" as opposed to peripheral information from proprioceptive sources in the (real) limbs. Finally, although the operation of the coupled action system involves regulating $\underset{\sim}{\Delta\phi}(=-\underset{\sim}{\Delta\Theta})$ and

$\underline{\alpha}\dot{\underline{\phi}}(=-\underline{\alpha}\dot{\underline{\theta}})$ to be "small," this is not solely a function of $\underline{\theta}$-position control requirements per se but also serves to validate the "small" relative displacement and velocity assumptions used for the real arm control matrices specified in relationship (29).

In summary, the network coupling version of task dynamics may provide a more biologically relevant sensorimotor control scheme than does the control law version. For single degree of freedom positioning tasks, it provides a rational account of the centrally specified rest angle's trajectory for these tasks without needing to invoke an explicitly preplanned representation of that trajectory. Rather, the rest angle trajectory evolves, even in the deafferented case, as an ongoing function of the underlying task dynamics. Similarly, when applied to discrete planar reaching tasks of 2-joint arms, the rest configuration trajectory will evolve so that the hand should move in a quasi-straight-line from initial to final position. Finally, when applied to cyclic spatial movements of a multijoint arm, the network coupling approach shares certain features with recent work on locomotion (cf. Grillner, 1981, for review). Investigators in this field assume the existence of innate, endogenous, cellular networks that are: a) capable of driving the limbs according to the locomotor task without requiring peripheral information; yet b) can be modulated--in phase dependent ways--by this same peripheral input (e.g., Forssberg, Grillner, & Rossignol, 1975). Task networks may be interpreted as the abstract, learned analogs of such concretely defined, innate networks. Thus, from a task dynamic perspective, the origins of task networks lie in the active discovery and specification processes that occur during skill learning. Once acquired, their operation is tailored to (tuned by) currently perceived task demands and the actor-environment spatial context.

References

Abbs, J. H., & Gracco, V. L. (in press). Control of complex motor gestures: Orofacial muscle responses to load perturbations of the lip during speech. Journal of Neurophysiology.

Abraham, R. H., & Shaw, C. D. (1982). Dynamics--The geometry of behavior. Santa Cruz, CA: Aerial Press.

Benati, M., Gaglio, S., Morasso, P., Tagliasco, V., & Zaccaria, R. (1980). Anthropomorphic robotics. I. Representing mechanical complexity. Biological Cybernetics, 38, 125-140.

Bizzi, E. (1980). Central and peripheral mechanisms in motor control. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Bizzi, E., & Abend, W. (1982). Posture control and trajectory formation in single and multiple joint arm movements. In J. E. Desmedt (Ed.), Brain and spinal mechanisms of movement control in man: New developments and clinical applications. New York: Raven Press.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1981). Processes underlying arm trajectory formation. In C. Ajmone-Marsan & O. Pompeiano (Eds.), Brain mechanisms of perceptual awareness and purposeful behavior (IBRO Monograph Series, pp. 311-318). New York: Raven Press.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1982). Arm trajectory formation in monkeys. Experimental Brain Research, 46, 139-143.

Bizzi, E., Chapple, W., & Hogan, N. (1982). Mechanical properties of muscles: Implications for motor control. Trends in Neurosciences, 5, 395-398.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland.

Delatizky, J. (1982). Final position control in simulated planar horizontal arm movements. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science.

Dorf, R. C. (1974). Modern control systems (2nd ed.). Reading, MA: Addison-Wesley.

Evarts, E. V. (1971). Feedback and corollary discharge: A merging of the concepts. In Central control of movement. Neurosciences Research Program Bulletin, 9, 86-112.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks. Biophysics, 11, 766-775.

Fel'dman, A. G., & Latash, M. L. (1982). Interaction of afferent and efferent signals underlying joint position sense: Empirical and theoretical approaches. Journal of Motor Behavior, 14, 174-193.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. Journal of Speech and Hearing Research, 18, 207-220.

Forssberg, H., Grillner, S., & Rossignol, S. (1975). Phase dependent reflex reversal during walking in chronic spinal cats. Brain Research, 55, 247-304.

Fowler, C. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press.

Georgopoulos, A. P., Kalaska, J. F., & Massey, J. T. (1981). Spatial trajectories and reaction times of aimed movements: Effects of practice, uncertainty, and change in target location. Journal of Neurophysiology, 46, 725-743.

Greene, P. H. (1971). Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, MA: MIT Press.

Grillner, S. (1981). Control of locomotion in bipeds, tetrapods, and fish. In J. M. Brookhart & V. B. Mountcastle (Eds.), Handbook of physiology, section 1: The nervous system; vol. II: Motor control, part 1 (pp. 1179-1236). Bethesda, MD: American Physiological Society.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. Oxford: Pergamon Press.

Hebb, D. O. (1949). The organization of behavior. New York: Wiley.

Hogan, N. (1980). Mechanical impedance control in assistive devices and manipulators. In Proceedings of the Joint Automatic Control Conferences, San Francisco, Vol. 1, TA-10B.

Hogan, N., & Cotter, S. L. (1982). Cartesian impedance control of a nonlinear manipulator. In W. J. Book (Ed.), Robotics research and advanced applications (pp. 121-128). New York: ASME.

50

Hollerbach, J. M. (1982). Computers, brains, and the control of movement. Trends in Neurosciences, 5, 189-192.

Hollerbach, J. M., & Flash, T. (1981). Dynamic interactions between limb segments during planar arm movement (AIM-635). Boston: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

Horak, F., & Nashner, L. (1983). Two distinct strategies for stance posture control: Adaptation to altered support surface configuration. Society of Neuroscience Abstracts, 9.

Ito, M. (1982). Questions in modeling the cerebellum. Journal of Theoretical Biology, 99, 81-86.

Jordan, D. W., & Smith, P. (1977). Nonlinear ordinary differential equations. Oxford: Clarendon Press.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 43, 1183-1196.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 49-70). New York: North-Holland.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.

Kelso, J. A. S., & Saltzman, E. L. (1982). Motor control: Which themes do we orchestrate? The Behavioral and Brain Sciences, 5, 554-557.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the coordination of two-handed movements. Journal of Experimental Psychology: Human Perception and Performance, 5, 229-238.

Kelso, J. A. S., Tuller, B., & Fowler, C. A. (1982). The functional specificity of articulatory control and coordination. Journal of the Acoustical Society of America, 72, S103.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. (1983). A 'dynamic pattern' perspective on the control and coordination of movement. In P. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag.

Klein, C. A., & Huang, C.-H. (1983). Review of pseudoinverse control for use with kinematically redundant manipulators. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13, 245-250.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Lashley, K. S. (1930). Basic neural mechanisms in behavior. Psychological Review, 37, 1-24.

Litvintsev, A. I. (1972). Vertical posture control mechanisms in man. Automation and Remote Control, 33, 590-600.

Mason, M. T. (1981). Compliance and force control for computer controlled manipulators. IEEE Transactions on Systems, Man, and Cybernetics, SMC-11, 418-432.

McGhee, R. B., & Iswandhi, G. I. (1979). Adaptive locomotion of a multi-legged robot over rough terrain. IEEE Transactions on Systems, Man, and Cybernetics, SMC-9, 176-182.

Minorsky, N. (1962). Nonlinear oscillations. Princeton, NJ: Van Nostrand.

Morasso, P. (1981). Spatial control of arm movements. Experimental Brain Research, 42, 223-227.

Nashner, L. M. (1979). Organization and programming of motor activity during posture control. In R. Granit & O. Pompeiano (Eds.), Reflex control of posture and movement (Progress in Brain Research, Vol. 50, pp. 177-184). New York: Elsevier/North-Holland Biomedical Press.

Nashner, L. M. (1981). Analysis of stance posture in humans. In A. L. Towe & E. S. Luschei (Eds.), Handbook of behavioral neurobiology: Vol. 5: Motor coordination (pp. 527-565). New York: Plenum.

Nashner, L. M., & Woolacott, M. (1979). The organization of rapid postural adjustments of standing humans: An experimental-conceptual model. In R. E. Talbott & D. R. Humphrey (Eds.), Posture and movement (pp. 243-257). New York: Raven Press.

Nashner, L. M., Woolacott, M., & Tuma, G. (1979). Organization of rapid responses to postural and locomotor-like perturbations of standing man. Experimental Brain Research, 36, 463-476.

Pellionisz, A., & Llinas, R. (1979). Brain modeling by tensor network theory and computer simulation. The cerebellum: Distributed processor for predictive coordination. Neuroscience, 4, 323-348.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. Science, 201, 1235-1237.

Raibert, M. H. (1978). A model for sensorimotor control and learning. Biological Cybernetics, 29, 29-36.

Raibert, M. H., Brown, H. B. Jr., Chepponis, M., Hastings, E., Shreve, S. E., & Wimberly, F. C. (1981). Dynamically stable legged locomotion. (Technical Report CMU-RI-TR-81-9). Pittsburgh: Carnegie Mellon University, The Robotics Institute.

Raibert, M. T., & Craig, J. J. (1981). Hybrid position/force control of manipulators. ASME Journal of Dynamic Systems, Measurement, and Control, 102, 126-133.

Rand, R. H., & Holmes, P. J. (1980). Bifurcation of periodic motions in two weakly coupled van der Pol oscillators. International Journal of Non-Linear Mechanics, 15, 387-399.

Sakitt, B. (1980). A spring model and equivalent neural network for arm posture control. Biological Cybernetics, 37, 227-234.

Saltzman, E. (1979). Levels of sensorimotor representation. Journal of Mathematical Psychology, 20, 91-163.

Schmidt, R. A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics.

Schmidt, R. A., & McGown, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. Journal of Motor Behavior, 12, 149-161.

Soechting, J. F. (1982). Does position sense at the elbow joint reflect a sense of elbow joint angle or one of limb orientation? Brain Research, 248, 392-395.

Soechting, J. E., & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. Journal of Neuroscience, 1, 710-720.

Stein, R. B. (1982). What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, 5, 535-577.

Szentagothai, J., & Arbib, M. A. (Eds.). (1974). Conceptual models of neural organization. Neurosciences Research Program Bulletin, 12(3).

Turvey, M. T., & Shaw, R. E. (1979). The primacy of perceiving: An ecological reformulation of perception for understanding memory. In L-G. Nilsson (Ed.), Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary. Hillsdale, NJ: Erlbaum.

Turvey, M. T., Shaw, R. E., & Mace, W. (1978). Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), Attention and performance VII. Hillsdale, NJ: Erlbaum.

Viviani, P., & Terzuolo, V. (1980). Space-time invariance in learned motor skills. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Whitney, D. E. (1972). The mathematics of coordinated control of prosthetic arms and manipulators. ASME Journal of Dynamic Systems, Measurement and Control, 94, 303-309.

Appendix A (Equation 7)

The body spatial variables ($\underset{\sim}{x}$, $\dot{\underset{\sim}{x}}$, $\ddot{\underset{\sim}{x}}$) of equation (6) are transformed into the joint variables ($\underset{\sim}{\phi}$, $\dot{\underset{\sim}{\phi}}$, $\ddot{\underset{\sim}{\phi}}$) or a massless arm model using the following kinematic relationships:

$$\underset{\sim}{x} = \underset{\sim}{x}(\underset{\sim}{\phi}) \qquad \text{(A1)}$$

$$\dot{\underset{\sim}{x}} = J(\underset{\sim}{\phi})\dot{\underset{\sim}{\phi}} \qquad \text{(A2)}$$

$$\ddot{\underset{\sim}{x}} = J(\underset{\sim}{\phi})\ddot{\underset{\sim}{\phi}} + (dJ(\underset{\sim}{\phi})/dt)\dot{\underset{\sim}{\phi}} \qquad \text{(A3)}$$

$$= J(\underset{\sim}{\phi})\ddot{\underset{\sim}{\phi}} + V(\underset{\sim}{\phi})\dot{\underset{\sim}{\phi}}_p, \text{ where}$$

$\underset{\sim}{x}(\underset{\sim}{\phi})$ = the current body spatial position vector of the terminal device expressed as a function of the current model arm configuration;

$$= (l_1 \sin\phi_1 + l_2 \sin(\phi_1 + \phi_2),\ -l_1\cos\phi_1 - l_2\cos(\phi_1 + \phi_2))^T;$$

$\dot{\underset{\sim}{\phi}}_p = [\dot{\phi}_1^2,\ \dot{\phi}_1\dot{\phi}_2,\ \dot{\phi}_2^2]^T$, the current joint velocity product vector;

$J(\underset{\sim}{\phi})$ = the Jacobian transform matrix;

$$= \begin{bmatrix} (l_1\cos\phi_1 + l_2\cos(\phi_1 + \phi_2) & l_2\cos(\phi_1 + \phi_2) \\ (l_1\sin\phi_1 + l_2\sin(\phi_1 + \phi_2) & l_2\sin(\phi_1 + \phi_2) \end{bmatrix};$$

$V(\underset{\sim}{\phi})$ = a matrix resulting from rearranging the terms of the expression $(dJ(\underset{\sim}{\phi})/dt)\dot{\underset{\sim}{\phi}}$ in order to segregate the joint velocity products into a single vector $\dot{\underset{\sim}{\phi}}_p$;

$$= \begin{bmatrix} (-l_1\sin\phi_1 - l_2\sin(\phi_1 + \phi_2)) & -2l_2\sin(\phi_1 + \phi_2) & -l_2\sin(\phi_1 + \phi_2) \\ (\ l_1\cos\phi_1 + l_2\cos(\phi_1 + \phi_2) & 2l_2\cos(\phi_1 + \phi_2) & l_2\cos(\phi_1 + \phi_2) \end{bmatrix}.$$

Making these substitutions into (6) and rearranging, we get equation (7):

$$M_B J\ \ddot{\underset{\sim}{\phi}} + B_B J\dot{\underset{\sim}{\phi}} + K_B \Delta\underset{\sim}{x}(\underset{\sim}{\phi}) = -M_B V\dot{\underset{\sim}{\phi}}_p, \qquad \text{(A4),(7)}$$

It should be noted that since $\Delta\underset{\sim}{x}$ in equation (6) is <u>not</u> assumed "small," the differential approximation $d\underset{\sim}{x} = J(\underset{\sim}{\phi})d\underset{\sim}{\phi}$ is not justified and, therefore, equation (A1) was used instead for the kinematic displacement transformation into model arm variables.

Appendix B (Equation 9)

One may derive using a Lagrangian analysis (see, for example, Saltzman, 1979, for details) the passive mechanical equations of motion for the 2-segment arm (frictionless, no gravity) described in the text:

$$M_A\ddot{\underset{\sim}{\theta}} + S_A\dot{\underset{\sim}{\theta}}_p = \underset{\sim}{0}, \quad \text{where} \qquad (A1), \quad (9)$$

$M_A = M_A(\underset{\sim}{\theta})$, the 2x2 acceleration sensitivity matrix with elements q_{ij}, where

$$q_{11} = m_2(l_1^2 + (1/3)l_2^2 + l_1l_2\cos\theta_2) + m_1(1/3)l_1^2$$

$$q_{12} = m_2((1/3)l_2^2 + (1/2)l_1l_2\cos\theta_2)$$

$$q_{21} = q_{12}$$

$$q_{22} = (1/3)m_2l_2^2;$$

$S_A = S_A(\underset{\sim}{\theta})$, a 2x3 matrix with elements s_{ij} resulting from rearranging the terms of the coriolis and centripetal torque terms in order to segregate the joint velocity products into a single vector $\dot{\underset{\sim}{\theta}}_p$, where

$$s_{11} = 0;\ s_{12} = -m_2l_1l_2\sin\theta_2;\ s_{13} = (1/2)s_{12}$$

$$s_{21} = -s_{13};\ s_{22} = 0;\ s_{23} = 0$$

Appendix C (Equation 15)

I) K_A. We begin with the expression $M_A^{-1}K_A\Delta\underset{\sim}{\theta}|_{\underset{\sim}{\theta}_o=\underset{\sim}{\phi}}$ from equation (11). Since we assume that $\Delta\underset{\sim}{\theta}$ is "small," we are justified in making the differential approximation:

$$[M_A^{-1}K_Ad\underset{\sim}{\theta}]|_{\underset{\sim}{\theta}_o} = [M_A^{-1}K_AJ^{-1}d\underset{\sim}{x}]|_{\underset{\sim}{\theta}_o}, \qquad \text{where} \qquad (C1)$$

$d\underset{\sim}{x} = \underset{\sim}{x}(\underset{\sim}{\theta}) - \underset{\sim}{x}(\underset{\sim}{\theta}_o)$ denotes the differential body space displacement between the terminal devices of the real (articulator network) and model (task network) arms, and $[M_A^{-1}K_AJ^{-1}]|_{\underset{\sim}{\theta}_o}$ denotes the articulator stiffness pattern governing the real arms's responses to small displacements about $\underset{\sim}{x}(\underset{\sim}{\theta}_o = \underset{\sim}{\phi})$.

The body spatial stiffness responses of the model arm specified by task dynamics for (possibly) large scale displacements $\Delta\underset{\sim}{x}(\underset{\sim}{\phi}) = \underset{\sim}{x}(\underset{\sim}{\phi}) - \underset{\sim}{x}_o$ from the reaching target $\underset{\sim}{x}_o$ are governed by the spatial restoring force term $[J^{-1}M_B^{-1}K_B\Delta\underset{\sim}{x}(\underset{\sim}{\phi})]|_{\underset{\sim}{\phi}}$ in equation (8). Assuming that the model ($\underset{\sim}{\phi} = \underset{\sim}{\theta}_o$) and real ($\underset{\sim}{\theta}$) arm configurations are "close," we compare the stiffness expressions and define the following constraint relationship:

$$K_A = [M_AJ^{-1}M_B^{-1}K_BJ]|_{\underset{\sim}{\theta}_o=\underset{\sim}{\phi}} \qquad (C2)$$

This relationship specifies that stiffness responses of the real arm to small $\Delta\underset{\sim}{\theta}$ perturbations will be defined according to task space axis directions and task space stiffness weightings.

II) B_A, $\underset{\sim}{\tau}_{Aa}$. Assuming that both $\Delta\underset{\sim}{\theta}$ and $\Delta\dot{\underset{\sim}{\theta}}$ are "small," one may use equations (8) and (14b) to define the following constraint relationships:

$$B_A = [M_A J^{-1} M_B^{-1} B_B J]|_{\underset{\sim}{\theta}_o = \underset{\sim}{\theta}} \quad \text{, and} \qquad \text{(C3)}$$

$$\underset{\sim}{\tau}_{Aa} = [M_A J^{-1} \dot{V} - S_A]|_{\underset{\sim}{\theta}_o = \underset{\sim}{\theta}, \dot{\underset{\sim}{\theta}}_p} \qquad \text{(C4)}$$

Footnotes

[1]An effector system is the set of limb segments or speech organs used in a given action; a terminal device or end effector is the part of a controlled effector system that is directly related to the goal of a performed action. Thus, in a reaching task, the hand is the terminal device and the arm is the effector system; in a cup-to-mouth task, the grasped cup is the terminal device and the hand-arm system is the effector system; in a steady state vowel production task, the tongue body surface is the terminal device and the jaw-tongue system is the effector system.

[2]Different systems may have different types of escapements. For example, van der Pol and Rayleigh oscillators have related escapement terms that are continuous functions of the systems' states; the pendulum clock's escapement term is a discontinuous function of the system's state, injecting a pulse of energy at one or two discrete points in the cycle.

[3]For a task in which an arm is nonredundant, the number of controlled spatial variables for the terminal device is equal to the number of controlled joint angular variables for the arm. Hence the inverse kinematic transformation from spatial motions of the terminal device to corresponding arm joint angular motions is determinate. For a task in which the number of joint variables exceeds the number of spatial variables, this transformation is indeterminate and the arm is redundant. For redundant arms, one may specify the inverse kinematic transformation by: a) "freezing" the extra joints in the arm; b) adding extra controlled spatial variables to the task description; or c) specifying optimality criteria to be satisfied for the joint variables during the movement.

[4]Indeed, herein lies an important difference between the various versions of the mass-spring model (or equilibrium point hypothesis for discrete targeting behavior). In one widespread view that is restricted to single degree of freedom motions, muscles are represented by a pair of springs acting across a hinge in the agonist-antagonist configuration. The final equilibrium point is established by selecting a set of length-tension properties in opposing muscles (e.g., Bizzi, 1980; Cooke, 1980; Kelso, 1977). This view, at best, may work for deafferented muscle, but, as pointed out by Fel'dman and Latash (1982, p. 178) it is inadequate for muscles in natural conditions. Moreover, as we have taken pains to point out, it does not work for complex, multivariable tasks. An alternative view, which we elaborate upon here, is that the parallel between a single muscle and a spring is not a literal one. Instead, the mass-spring model is better viewed as a model of equifinality or motor equivalence: it is this abstract functional property that particular behaviors share with a mass-spring system (Kelso, Holt, Kugler, & Turvey, 1980; Kelso & Saltzman, 1982). In short, the former, articulator dynamic version is a hypothesis about a physiological mechanism whose shortcomings have been noted (Bizzi, Accornero, Chapple, & Hogan, 1982; Fel'dman & Latash, 1982). The latter, abstract dynamic version refers to a complex system, and is a hypothesis about behavioral function.

[5]Currently, our task-dynamic formulation does not include precision force control tasks. It can be easily adapted for tasks that demand particular motion patterns along a surface and only approximate control of the force exerted by the terminal device normal to the surface (e.g., polishing a car, erasing a blackboard). The approach can also be adapted for precision force control tasks, however, as demonstrated by Hogan and Cotter (1982).

[6]Mason (1981; see also Raibert & Craig, 1981) has formalized a related geometrical description for manipulator contact tasks in which different tasks are characterized by distinct generalized surfaces in a constraint space. In this task-specific constraint space, the task degrees of freedom are partitioned into those associated with either position or contact force control, respectively, during performances of the associated task. Such an approach requires, however, explicit task- and context-specific position and force trajectory plans for the task's terminal device. In contrast, the task dynamic approach requires no such explicit trajectory plans, due to the task-specific dynamical topologies defined for the task-space degrees of freedom. In our formulation, then, task-appropriate terminal device trajectories are emergent properties implicit in the corresponding underlying task-dynamic organizations.

[7]In redundant task-articulator situations (see footnote 3), J^{-1} is not defined and the Jacobian pseudoinverse (J^{+}) or weighted Jacobian pseudoinverse (J^{*}) may be used (Benati et al., 1980; Klein & Huang, 1983; Whitney, 1972). Using J^{*} provides an optimal weighted least squares solution for the differential transformation from spatial to joint motion variables. If this weighting is task dependent, then J^{*} would be both task- and configuration-dependent. For example, if a three-joint arm is used to position the fingertip in a spatially planar reaching task, different weightings would correspond to different arm joint motion strategies. One weighting might correspond to a predominantly shoulder motion strategy, while a second weighting might specify a predominantly elbow motion strategy, etc. In such cases, elements of the weighting matrices used for the corresponding weighted Jacobian pseudoinverses define a further set of tuning parameters for the task network.

[8]As we demonstrate via simulation (in the Trajectory Shaping section) in the case of our task space point attractor reaching example, it may be possible to ignore the velocity product torque terms, and therefore omit τ_{Aa} from equation (12), yet still arrive at the desired target via quasi-straight line hand trajectories. In fact, reach trajectories generated without such correction appear more similar to experimentally observed trajectories than ones generated with "perfect" velocity product torque correction.

[9]The desirability of using such scaling coefficients was pointed out by Mason (1981). In addition to using them to ensure dimensional homogeneity, Mason showed that different values could be used to provide correspondingly different weightings of rotational vs. linear aspects of task performances. However, since the task dynamic approach uses relative task axis stiffness weightings for this purpose, the value of ρ was simply set to 1.0 in our treatments.

[10]For task spaces not defined by point attractors along each task axis, however, equation 29a will no longer hold. For example, if a given task has a limit cycle organization for one task axis, and therefore a nonlinear damping term, the B_T and hence B_B matrices will reflect only the linear negative part of this damping. If B_B were used in equation 29a, the articulator network should be highly unstable. In such cases, however, one might simply choose B_A to make the articulator network stable about $\underset{\sim}{\theta}_o$, given the K_A specified in equation 29b.

SPECULATIONS ON THE CONTROL OF FUNDAMENTAL FREQUENCY DECLINATION*

Carole E. Gelfer,+ Katherine S. Harris,+ Rene Collier,++ and Thomas Baer

Introduction

It is generally assumed that, for read speech at least, the fundamental frequency of the voice declines over the course of major syntactic constituents. These units correspond to what has previously been termed the "breath group" (Lieberman, 1967; Lieberman, Sawashima, Harris, & Gay, 1970) or "intonation group" (Breckenridge, 1977), being marked on either end by a pause and/or inspiration. The general downdrift of F_0 is exclusive of local perturbations secondary to syllable prominence and segmental effects, and is probably best characterized by a steadily declining baseline upon which these local movements are superimposed (Cohen, Collier, & t'Hart, 1982; Fujisaki & Hirose, 1982).

Variations in subglottal pressure (P_s) and cricothyroid (CT) muscle activity are thought to bear most directly on F_0 variation, although it has been difficult to separate the CT's contribution to the global prosodic structure of an utterance from its involvement in ongoing local adjustments. However, despite these methodological problems, there has been little evidence to suggest a gradual decline in CT activity corresponding to that in F_0. Rather, the CT's most active involvement in intonation appears to be confined to instances of local emphasis (e.g., Collier, 1975; Maeda, 1976). Subglottal pressure, on the other hand, does exhibit a declination of its own that at least grossly mirrors the F_0 contour (Atkinson, 1973; Collier, 1975; Lieberman, 1967; Maeda, 1976), thus suggesting that F_0 declination might be a passive phenomenon. However, despite the apparent relationship between P_s and F_0, attempts to establish a direct correlation between the two (Atkinson, 1973; Maeda, 1976) have been unsuccessful in that the drop in F_0 exceeds the 3-7 Hz/cm-H_2O that a purely passive model would predict (Baer, 1979; Hixon, Klatt, & Mead, 1971; Ladefoged, 1963).

Some researchers have proposed that declination, and the physiological processes underlying it, is under active speaker control. This assumption derives in part from observations of variations in some aspects of F_0 as a func-

---

*To appear in Vocal Fold Physiology: Physiology and Biophysics of Voice. Proceedings of a Conference, University of Iowa, May 4, 1983. A version of the paper was presented at the 105th Meeting of the Acoustical Society of America, Cincinnati, Ohio, May 9-13, 1983.

+Also Graduate School, City University of New York.

++University of Antwerp.

Acknowledgment. This work was supported by NINCDS Grants NS-13870 and NS-13617 to Haskins Laboratories. We thank J. A. S. Kelso for his helpful comments and suggestions.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

tion of utterance length. Cooper and Sorenson (1981), for example, found significant, if not robust, increases in initial peak F_0 for progressively longer utterances, while Breckenridge (1977) and Maeda (1976) observed the total amount of declination to be relatively constant under the same conditions. However, there is a large amount of data to suggest that final F_0 values are invariant despite changes in the length of utterances (Boyce & Menn, 1979; Cooper & Sorenson, 1981; Kutik, Cooper, & Boyce, 1983; Maeda, 1976), initial starting frequency (Liberman & Pierrehumbert, 1982), or the insertion of dependent clauses such as parentheticals (Kutik et al., 1983). What these results seem to suggest, then, is that, as an utterance increases in length, either the total amount of declination increases or the rate of decline decreases. However, it is not entirely clear whether 1) these are mutually exclusive aspects of F_0 declination and 2) length-dependent variations in F_0 necessarily refute the predictions of a passive model of declination and favor theories involving elaborate speaker pre-planning.

The present study examined the F_0 declination, and some physiological variables presumed to underlie it, under various linguistic conditions. Our purpose was to elucidate further the relationships among these variables and to speculate whether speakers exercise significant control over any or all of them.

Methods

The subject was a native speaker of Dutch who produced five repetitions of Dutch utterances of three lengths; six, thirteen, and twenty syllables. Mean utterance durations were 1.35, 2.065, and 3.02 seconds, respectively. All three lengths had the first four syllables in common; for the longer utterances, the first eight syllablies were identical (see Appendix). Each utterance type[2] was also produced in reiterant form, using either the syllable /ma/ or /fa/.[2] The purpose of employing reiterant speech was to neutralize segmental effects while preserving overall intonation and syllable timing (Larkey, 1983; Liberman & Streeter, 1978). In addition, by using syllables with expected differences in airflow requirements, the effect of these differences on subglottal pressure and, possibly, F_0, could be assessed.

For each length condition, emphatic stress was placed either on the first syllable receiving lexical stress (the second syllable in the utterance), the last syllable receiving lexical stress (the penultimate syllable) or both. We will refer to these as early, late, and double stress conditions, respectively. In all, there were twenty-seven utterance types (3 phonetic conditions x 3 stress conditions x 3 length conditions). All tokens were aligned to the onset of the second vowel and averaged for each utterance.

The results were analyzed with respect to the effects of utterance length and syllable emphasis on initial F_0, P_s, CT, and respiratory activity, and the magnitude and rate of decline in each of these variables over entire utterances.

Subglottal pressure was recorded by means of a pressure transducer inserted through the cricothyroid membrane into the trachea. Standard EMG techniques were used to record from the cricothyroid muscle (Harris, 1981). Lung volume was inferred from the calibrated sum of thoracic and abdominal signals from a Respitrace inductive plethysmograph, and F_0 was derived from

the output of an accelerometer attached to the pretracheal skin surface. A cepstral technique was used to extract F_0 from the signal.

Results

Figure 1 shows Respitrace comparisons for each phonetic condition across stress types for Length 2 utterances. Within utterances of a given phonetic composition (i.e., Dutch, /ma/ or /fa/), the rate of air expenditure appears to remain constant within stress condition, as is obvious from the generally parallel tracings. However, the peak inspiration varies inconsistently across parallel sets.[3] Thus, it would seem that, on the respiratory level, local variables such as the degree or place of emphasis were not reflected in the air flow management of this speaker's utterances.

Across phonetic conditions, however, airflow rates do differ, as is evidenced by the apparent differences in the rate at which these curves decline. The left-hand section of Figure 2 shows a comparison of the Respitrace curves for each phonetic condition for early stress across three utterance lengths. It appears that airflow rate for the /fa/ condition always exceeds that for the /ma/ condition, while, for the Dutch, air expenditure is more variable. The obvious question is whether these differences in airflow are reflected in the pressure. From the corresponding subglottal pressure tracings on the right of Figure 2, it can be seen that they are not. Furthermore, while local segmental effects are apparent in the P_s curves, particularly for the Dutch utterances, it is also apparent that, for the three comparisons made at each length, a single line could characterize the decline of subglottal pressure, despite the variations in phonetic composition and concomitant airflow characteristics.

Because of the demonstrated uniformity of P_s across phonetic conditions, the remainder of this paper will focus on the analysis of the reiterant /ma/ utterances on the assumption that they are at least generally representative of normal speech.

Table 1

Peak inspiration (left) and total inspiration (right), in liters, for the three length conditions across all stress types.

	Peak Inspiration (liters)				Amount Inspiration (liters)			
	Early	Double	Late	Mean	Early	Double	Late	Mean
Length 1	3.83	4.06	4.05	3.98	.85	.93	1.17	.98
Length 2	4.1	4.35	4.12	4.19	.99	1.36	1.11	1.15
Length 3	4.23	4.16	4.09	4.16	1.41	1.61	1.26	1.43
Mean	4.05	4.19	4.09		1.08	1.3	1.18	

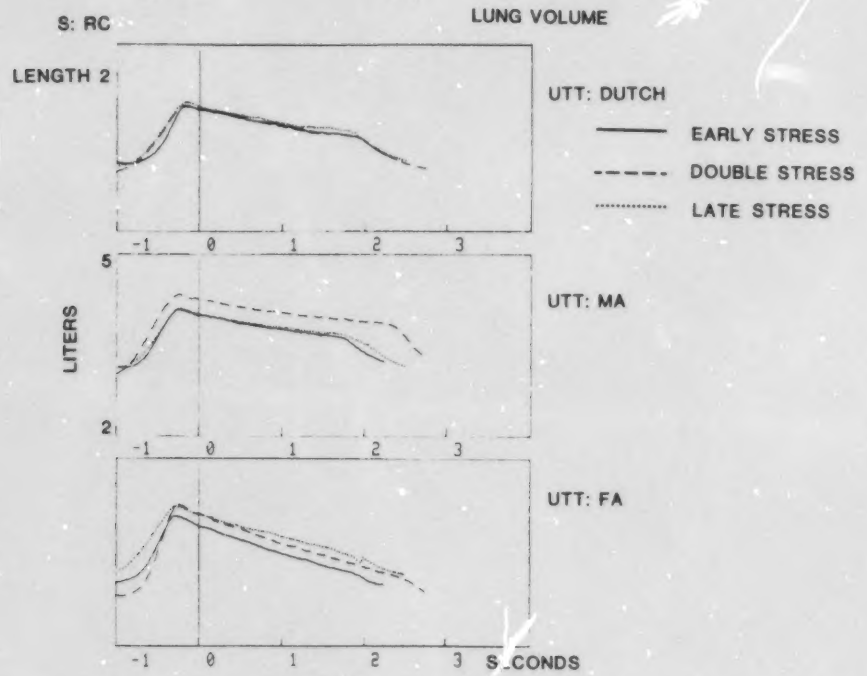


Figure 1. Comparison of Respitrace curves for Length 2 utterances across all stress types shown for each of the three phonetic conditions.

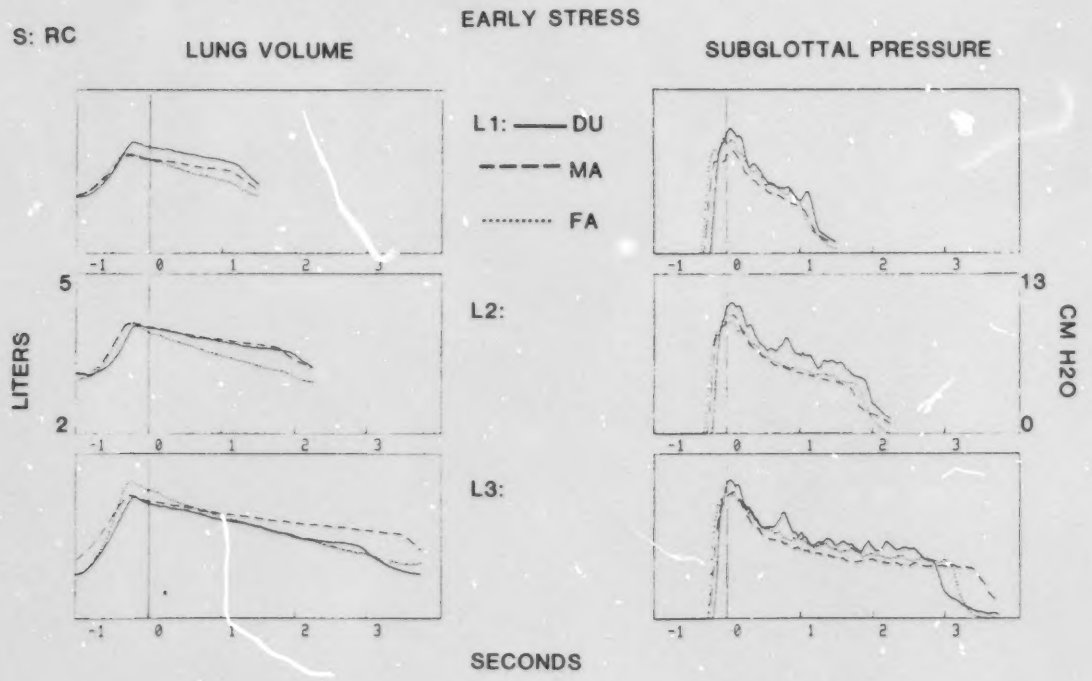


Figure 2. Corresponding Respitrace (left) and subglottal pressure (right) curves for the early stress condition across the three phonetic conditions. Comparisons are shown for each utterance length.

Figure 3 again shows Respitrace curves for each stress type across the three utterance lengths. It can be seen that there are no visually significant differences in the rate of expiration nor evidence of systematic adjustments in peak inspiration as a function of anticipated length. It is the case, though, that the depth of inspiration (with the exception of one utterance type) appears to be adjusted according to utterance length. This is evident from the values in Table 1, which shows both the point of peak inspiration and the amount of inspiration (calculated by subtracting the preceding valley from the peak) for each utterance. However, because the experiment was designed in such a way that all tokens of a particular stress type were produced in blocks of utterances of increasing length, it is impossible to determine the significance of this finding. In other words, because Length 3 tokens were always preceded by tokens of the same length or, in one instance, by the last token of the shorter length utterance, inspiration necessarily began at a point lower in this speaker's vital capacity than, for example, Length 2 tokens, which could have only been preceded by tokens of the same length or shorter. Thus, we are unable to determine whether the increase in the depth of inspiration as a function of length represents an artifact of experimental design or evidence of anticipated pulmonary requirements. Overall, the Respitrace data fail to demonstrate conclusively the manner or extent to which this speaker makes prephonatory adjustments of this kind under the various conditions. However, in light of the otherwise uniform nature of these Respitrace curves, and the absence of any obvious relationship with the subglottal pressure, their influence on the ultimate trajectory of fundamental frequency declination appears questionable.

Figure 4 depicts the F_0 contours for the three stress conditions for each utterance length. These contours probably represent what has been termed "baseline declination" in as pure a form as possible in that significant segmental effects are absent. For the early and double stress conditions, there is an obvious peak associated with every emphatic syllable, and a consistent initial peak height difference as a function of utterance length. However, F_0 does not decline steadily from these peaks. Rather, there is a rapid drop in frequency to a point from which F_0 then begins a steady decline. While the time course of this initial plunge is constant across lengths, despite differences in peak height, the points from which the slow decline begins for each length are not, bearing instead the same relationship as the initial peaks. This relationship appears to be maintained throughout the course of at least the longer utterances, although they appear to decline in parallel. In the absence of early emphasis in the late stress condition, the F_0 peaks occur upon initiation of the utterance and are thus displaced in time relative to the second syllable peaks in the former two conditions. Furthermore, the decline of F_0 from these peaks is far more gradual and less strikingly parallel. However, it is of some interest to note that the relationship of these nonemphatic initial peaks across lengths is the same as for their emphatic counterparts.

Figure 5 shows the corresponding subglottal pressure tracings. It can be seen that the same general tendencies prevail. That is, there is an effect of utterance length on the initial peak pressure and a relatively rapid initial pressure drop into a more-or-less parallel and steadily declining function for the longer utterances. Again, the peaks occur earlier in the late stress utterances and the initial pressure drop is less rapid.

If the F_0 and P_s tracings are examined in parallel, it becomes apparent that there is a point in time, following the initial peaks, after which the decline in F_0 almost mirrors that of P_s. However, the parallelism is less

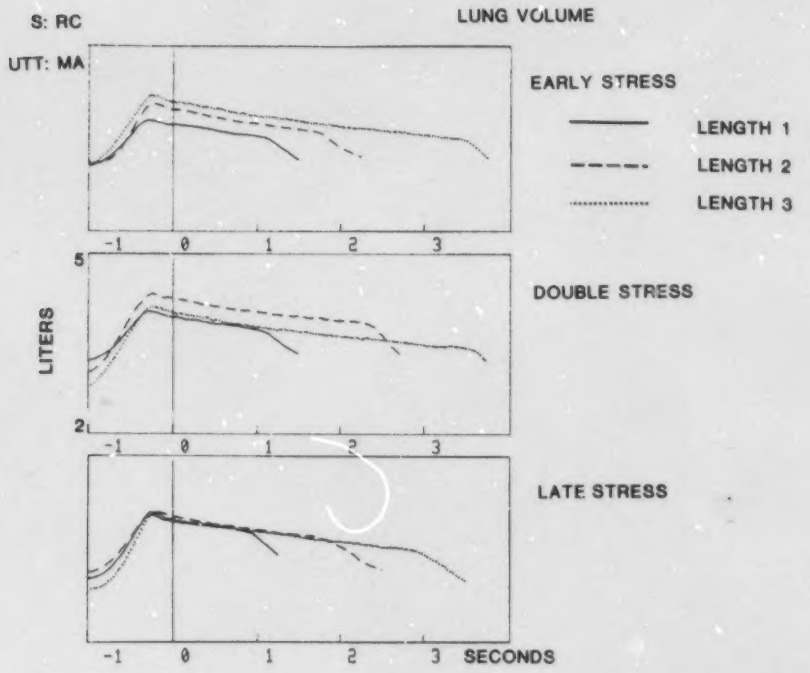


Figure 3. Respitrace curves for reiterant /ma/ utterances across lengths. Comparisons are shown for each stress condition.

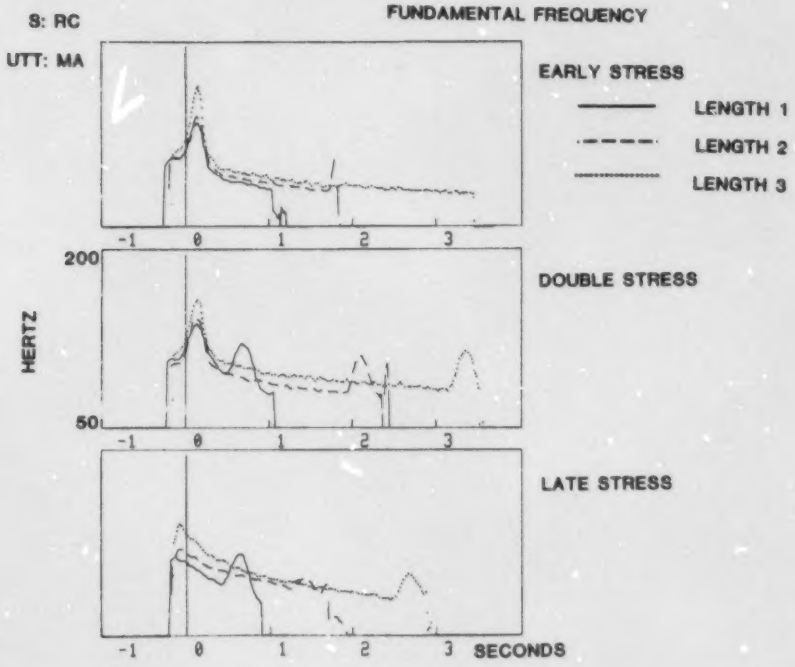


Figure 4. Fundamental frequency curves for reiterant /ma/ utterances across lengths. Comparisons are shown for each stress condition.

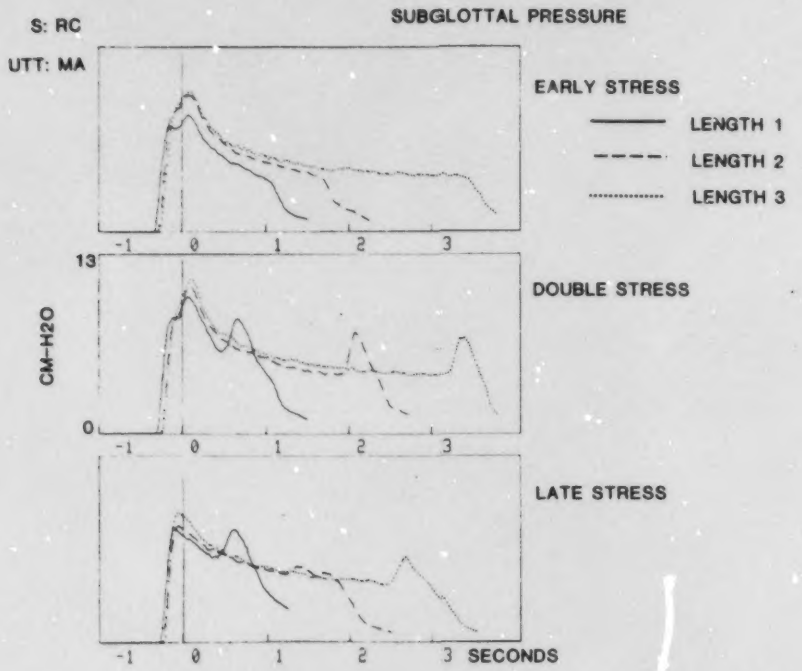


Figure 5. Subglottal pressure curves for reiterant /ma/ utterances across lengths. Comparisons are shown for each stress condition.

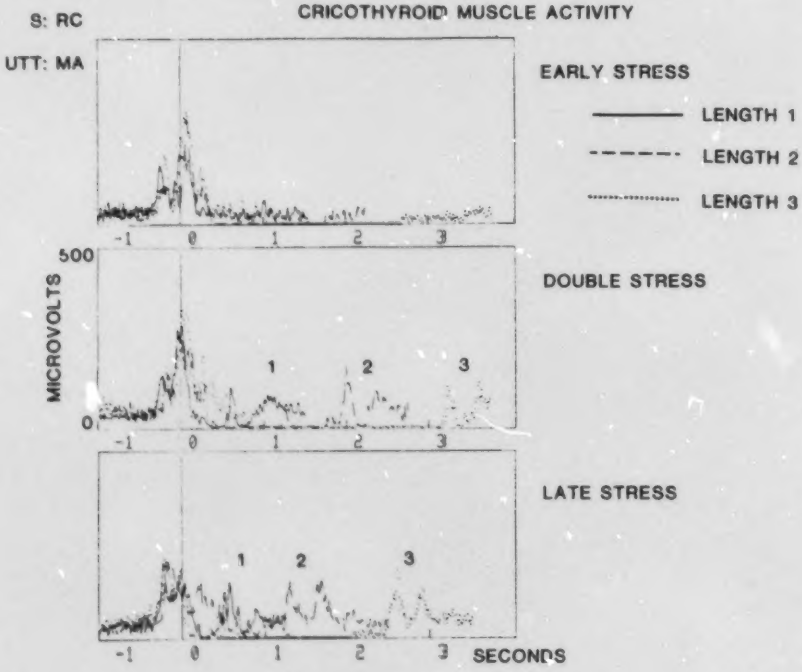


Figure 6. Averaged Cricothyroid muscle activity for reiterant /ma/ utterances across lengths for each stress condition. Final peaks for each utterance length are denoted by the numbers above these peaks.

obvious for the earlier portions of early and double stress utterances, since the most rapid drop in P_s is far more gradual than that for F_0.

Figure 6 shows CT activity across lengths for the three stress conditions. The overall CT pattern differs from those of P_s and F_0 in that, although inherently noisy, CT activity appears to be relatively binary in utterances of this form. There are significant increases in CT occurring for initial syllables, whether stressed or not, and for all final stressed syllables, which in this figure are marked in the double and late stress conditions according to their respective lengths. CT activity during the early portion of these utterance is characterized by double peaks whose timing is identical across stress types, but whose relative magnitudes differ with stress type, corresponding to the placement of the P_s and F_0 peaks. The double peaks associated with the final stressed syllables of Lengths 2 and 3, however, are the result of averaging events that are distant from the line-up point in tokens of slightly unequal lengths, and are not characteristic of CT activity for final stress peaks.

In order to examine the effect of anticipated length on the initial portions of utterances, we compared initial peak values of CT, P_s, and F_0 for each stress type across lengths. It should be recalled that the initial utterance peaks for P_s and F_0 in the late stress condition were displaced relative to those with early and double stress, while the timing of the CT peaks remained constant irrespective of stress type. In the interest of consistency, then, the values reported here for P_s and F_0 in the late stress condition are those that correspond in time to the peaks for the other two stress conditions and, thus, actually represent values on the declining portion of these curves. The results are shown in Table 2. It can be seen that a consistent effect of sentence length obtains for every stress condition for all physiological and acoustic measures.[4]

If the corresponding F_0 and CT curves are examined in parallel, there appears to be a close correspondence between the time course of the CT suppression and the point at which F_0 begins its steadiest decline. We would thus hypothesize that the combined activity of CT and P_s accounts for the behavior of F_0 near peaks, but not during the period of F_0 slow decline. We acknowledge, of course, that the activity of a number of muscles, not monitored in this study, may also have causal effects on F_0.

Assuming, then, that CT plays little or no active role in F_0 declination, we examined the relationship between P_s and F_0 in two different ways. First, the amount of drop in F_0 and P_s was calculated between the point at which the CT activity ceased and the end of the utterance in the early stress condition, and between CT cessation and the minimum values just preceding the last peak in the double and late stress conditions. In the second analysis, we used the average duration of Length 1 of the early stress utterances as a fixed endpoint and determined the amount and rate of F_0 and P_s decline between the offset of CT activity and this fixed endpoint for all utterances.

The offset of CT activity was defined as the time at which the EMG output (measured in microvolts) dropped to and remained below a level equivalent to the baseline plus 10% of the peak level. These analyses were not performed on Length 1 of the double and late stress conditions. In the former condition, the interval between CT offset for the first peak and CT onset for the second was too short. In the latter condition, CT activity was never consistently

Table 2

Initial peak measurements of cricothyroid activity, subglottal pressure, and fundamental frequency for the three length conditions across stress types. P_s and F_0 values for the late stress condition do not represent absolute peak values (see text).

Initial Peak Values

		Early	Double	Late	Mean	
CT	Length 1	202	273	159	211	Cricothyroid (μV)
	Length 2	296	277	169	247	
	Length 3	310	331	189	277	
	Mean	269	294	172		
P_s	Length 1	8.3	9.9	7.1	8.4	Subglottal Pressure ($cm-H_2O$)
	Length 2	9.7	10.7	7.4	9.3	
	Length 3	9.9	11.3	8.2	9.8	
	Mean	9.3	10.6	7.6		
F_0	Length 1	135	137	102	125	Fundamental Frequency (Hz)
	Length 2	141	143	111	132	
	Length 3	166	158	124	149	
	Mean	147	146	112		

Table 3

Analyses of rate of decline in F_0 and P_s across lengths for each stress condition, calculated for (1) the interval from the point of CT offset to P_s minima (variable interval) and (2) the interval from the point of CT offset to a fixed endpoint corresponding to the average duration of Length 1 of the Early stress condition (constant interval). The frequency-to-pressure ratios are also shown for each analysis.

		ANALYSIS 1			ANALYSIS 2		
		F0	Ps	F0/Ps	F0	Ps	F0/Ps
Early	Length 1	22.21	3.94	5.64	22.21	3.94	5.64
	Length 2	14.39	2.47	5.83	19.7	3.73	5.28
	Length 3	7.03	1.07	6.57	17.42	3.95	5.2
Double	Length 1	-	-	-	-	-	-
	Length 2	15.37	2.38	6.46	22.52	4.12	5.47
	Length 3	10.76	1.36	7.91	19.75	3.49	5.66
Late	Length 1	-	-	-	-	-	-
	Length 2	20.79	2.57	8.09	17.32	2.61	6.64
	Length 3	16.85	1.56	10.8	35.22	3.52	10.01

suppressed, so that an offset time could not be obtained. Furthermore, in both cases the designated interval for the second analysis extended into the final stress peak. These analyses were performed on a token-by-token basis in order to accommodate variability in the timing of CT activity.

Table 3 shows the results of both analyses in terms of F_0 slope (Hz/sec), P_s slope (cm-H_2O/sec), and the frequency-to-pressure ratio (Hz/cm-H_2O). Looking first at the ratios from both analyses, it should be noted that six of the seven values from Analysis 2 fall within the acceptable range of 3-7 Hz/cm-H_2O, while only four of the seven values from Analysis 1 fall within this range. However, even those values that fall outside the range are considerably lower than those reported when the effects of CT and possibly other muscle activity are not neutralized (see Maeda, 1976). Thus, a passive mechanism whereby F_0 declination is determined by a steadily falling subglottal pressure should be reconsidered.

Discussion

As for the influence of utterance length on the slope of F_0 and P_s change, the results of Analysis 1 show a substantial decrease in the rate of change with increasing length. This effect has been observed in previous studies and assumed to represent high level preplanning whereby certain physical aspects are represented in a speaker's utterance plan. However, when slope is calculated over fixed portions of these same utterances, as in Analysis 2, the length effect observed in Analysis 1 is substantially lessened, demonstrating a more constant rate of decline across lengths. (For Length 3 of the late stress condition, there is probably some peculiarity in the data, particularly for F_0.) The results of the latter analysis further suggest that neither F_0 nor P_s decline at a constant rate across an entire utterance. If they did, we would expect the slopes to be identical over any portion of a given utterance, despite its length. However, the results of Analysis 1 demonstrate that this is not the case. It appears that, with the obvious exception of the late stress utterances, the rate of decline in P_s and F_0 is greatest earlier in an utterance, as is indicated by the steeper slopes in the second analysis, and that these curves would be best characterized by an exponential function. Thus, the apparent "length effect" that we and others observe when slope is calculated over an entire utterance can probably be attributed to the nonlinear nature of F_0 declination and not to elaborate precalculations or ongoing reorganization on the basis of utterance to length. Our data substantiate the claims of Liberman and Pierrehumbert (1982) that the F_0 contour gradually approaches an asymptotic value, as well as Maeda's finding that the latter portion of some utterances may not show any evidence of declination.

The systematic adjustments in peak F_0 suggest that, on some level, this speaker does take sentence length into account. However, these peaks do not appear to influence the trajectory of the total declination contour. Rather, their influence appears to be limited to their immediate vicinity, probably including the frequency from which declination actually begins. However, the latter is probably a function of temporal constraints whereby, in a fixed amount of time, the frequency to which F_0 falls is a function of the frequency from which it starts. Thus, whatever its purpose, manipulating peak height does not appear to be essential to the realization of declination, per se.

In summary, we have found that, for reiterant utterances composed of voiced continuants, where normal segmental adjustments were presumed to be neutralized, CT activity was prominent in instances of emphatic syllable stress, and relatively inactive elsewhere. Subglottal pressure, on the other hand, showed a gradual decline before and/or after stress peaks and was paralleled by a falling fundamental frequency. Thus, while we cannot rule out effects such as vocal fold relaxation on F_0 and P_s during these intervals, the data do suggest that, where CT activity is negligible, F_0 declination can be accounted for on the basis of a falling P_s alone.

Our conclusions at this point must be tentative for two reasons: First, because we have anlayzed the data of only one subject and second, because there are inconsistencies between the late stress utterances and the other two stress conditions. However, we believe there are strong indications that declination may be the province of low-level processes such that variations in certain aspects of F_0 are the result, not of high-level (i.e. cognitively-generated) planning processes, but of the intrinsic behavioral properties of underlying physiological systems.

References

Atkinson, J. E. (1973). Aspects of intonation in speech: Implications from an experimental study of fundamental frequency. Unpublished doctoral dissertation, University of Connecticut.

Baer, T. (1979). Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. Journal of the Acoustical Society of America, 65, 1271-1275.

Boyce, S., & Menn, L. (1979). Peaks vary, endpoints don't: Implications for linguistic theory. Proceedings from the Fifth Annual Meeting of the Berkeley Linguistics Society.

Breckenridge, J. (1977). Declination as a phonological process. Murray Hill, NJ: Bell System Technical Memorandum.

Cohen, A., Collier, R., & t'Hart, J. (1982). Declination: Construct or intrinsic feature of pitch? Phonetica, 39, 254-273.

Collier, R. (1975). Physiological correlates of intonation patterns. Journal of the Acoustical Society of America, 58, 249-255.

Cooper, W. E., & Sorenson, J. M. (1981). Fundamental frequency in sentence production. New York: Springer-Verlag.

Fujisaki, H., & Hirose, K. (1982). Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. Preprints of papers, Working Group on Intonation (pp. 57-70). The XIIIth International Congress of Linguistics, Tokyo.

Harris, K. S. (1981). Electromyography as a technique for laryngeal investigation. In C. L. Ludlow & M. O. Hart (Eds.), ASHA Reports: Proceedings of the Conference on the Assessment of Vocal Pathology, 11, 70-86.

Hixon, T. J., Klatt, D. H., & Mead, J. (1971). Influence of forced transglottal pressure changes on vocal fundamental frequency. Journal of the Acoustical Society of America, 49, 105.

Kutik, E. J., Cooper, W. E., & Boyce, S. (1983). Declination of fundamental frequency in speakers' production of parenthetical and main clauses. Journal of the Acoustical Society of America, 73, 1723-1730.

Ladefoged, P. (1963). Some physiological parameters in speech. Language and Speech, 6, 109-119.

Larkey, L. S. (1983). Reiterant speech: An acoustic and perceptual evaluation. Journal of the Acoustical Society of America, 73, 1337-1345.

Liberman, M., & Pierrehumbert, J. (1982). Intonational invariance under changes in pitch range and length. Bell Laboratories Internal Technical Memorandum.

Liberman, M., & Streeter, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. Journal of the Acoustical Society of America, 63, 231-233.

Lieberman, P. (1967). Intonation, perception and language. Research Monograph No. 38. Cambridge: MIT Press.

Lieberman, P., Sawashima, M., Harris, K. S., & Gay, T. (1970). The articulatory implementation of the breath-group and prominence: Crico-thyroid muscular activity in intonation. Language, 46, 312-327.

Maeda, S. (1976). A characterization of American English intonation. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Footnotes

[1]This should not be interpreted as meaning that there is not a gradual relaxation of the muscle, but that the EMG activity associated with cricothyroid contraction does not appear to lessen gradually over the course of an utterance.

[2]The durations of the reiterant utterances are, on average, somewhat longer than the corresponding Dutch due to the intrinsic duration of /a/ and/or the inadvertent addition of extra syllables. Those tokens for which the latter was evident were still included in all analyses on the assumption that the speaker's intention was to produce an utterance of a given duration, so that any length-dependent adjustments would be identical.

[3]As used here, peak inspiration corresponds to the maximum amplitude of the output signal of the Respitrace immediately preceding the onset of speech. Because of the built-in filter characteristics of the Respitrace, however, this point may not represent actual peak inspiration. Furthermore, baseline drift and positional changes may introduce artifact into the signal as well. Therefore, peak inspiration measures should be interpeted with caution.

[4]The actual peak values for P_s and F_0 for the late stress condition evidence a similar length effect. They are, in order of increasing length: P_s: 7.9, 8.1, 9.0; F_0: 113, 119, 139.

70

APPENDIX

Early Stress

Length 1: Je weet dat jan nadenkt.
Length 2: Je weet dat jan erover nadenkt te betalen.
Length 3: Je weet dat jan erover nadenkt ons daaroor met genoegen te betalen.

Double Stress

Length 1: Je weet dat jan nadenkt.
Length 2: Je weet dat jan erover nadenkt te betalen.
Length 3: Je weet dat jan erover nadenkt ons daarvoor met genoegen te betalen.

Late Stress

Length 1: Je weet dat jan nadenkt.
Length 2: Je weet dat jan erover nadenkt te betalen.
Length 3: Je weet dat jan erover nadenkt ons daarvoor met genoegen te betalen.

SELECTIVE EFFECTS OF MASKING ON SPEECH AND NONSPEECH IN THE DUPLEX PERCEPTION PARADIGM

Shlomo Bentin+ and Virginia A. Mann++

Abstract. Perception of second formant transitions isolated from a synthetic [ba] or [ga] syllable as nonspeech "chirps" or as supporting identification of a stop consonant was investigated using the duplex perception phenomenon. This phenomenon arises when dichotic presentation of the transition and the remaining part of a CV syllable (the base) allows the transition to support both perception of that syllable and a nonspeech chirp simultaneously. Over the course of four experiments it was found that: 1) Stimulus onset asynchrony of base and transition impairs the accuracy of syllable labeling, but improves "chirp" classification into nonspeech categories, whereas a white noise backward mask ipsilateral with the transition impairs categorization of "chirps" but not of the syllables supported by these transitions. 2) Progressive attenuation of the relative intensity of the transitions impairs speech perception at a slower rate than nonspeech perception. 3) A white noise mask preceding the base in the ipsilateral ear and presented simultaneously with the transition interferes with labeling syllables, but does not affect categorization of chirps. 4) A white noise ipsilateral backward mask of the transition penalizes both categorization and discrimination of nonspeech percepts more extensively than that of speech. An analagous mask consisting of a second formant transition appropiate to [da] did not affect nonspeech perception, but impaired correct labeling of the syllables. It is suggested that perception in the speech and in the nonspeech modes is contingent upon activation of different central mechanisms.

Speech perception involves the recovery of phonetic information embedded in acoustic patterns that stimulate the auditory nervous system. Frequency-modulated acoustical signals, formant transitions, can be sufficient cues for the perceived distinction between stop consonants when they are integrated with a syllabic base, in which case they support an abstract phonetic percept

---

+Aranne Laboratory of Human Psychophysiology, Hadassah Hospital, Department of Neurology, Jerusalem.

++Also Bryn Mawr College.

Acknowledgment. This work was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories, and by Bryn Mawr College. In addition, S. Bentin was partly supported by the Lady Davis Foundation, awarded through the Hebrew University, Jerusalem. This study would not have been completed without the continuous encouragement and advice of Alvin M. Liberman. We also thank Bruno H. Repp for his valuable comments on previous versions of this manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

such as "b" or "g." However, when the same formant transitions are presented in isolation, perception reflects the acoustic characteristics of the time-varying nature of the frequency-modulated signal. Thus, although some subjects might be able to perceive isolated formant transitions as speech-like (Nusbaum, Schwab, & Sawusch, 1983), most tend to describe them as "chirps" that have no relation to the perceptual characteristics of the stop consonants that they otherwise may cue (Mattingly, Liberman, Syrdal, & Halwes, 1971). This radically different perception of the same physical stimulus in two different acoustic contexts has been taken to reflect two different modes of auditory perception in humans: a phonetic mode exclusively dedicated to speech, and a nonphonetic mode for the perception of other auditory stimuli (Liberman, 1982; Liberman & Studdert-Kennedy, 1978; Mann & Liberman, 1983; Repp, 1982).

The two different perceptual modes are simultaneously operative in the phenomenon of "duplex" perception first described by Rand (1974). Duplex perception occurs when the transition of the second and/or third formant, which supports perception of a unique stop consonant, is separated from the rest of a synthetic syllable and presented to one ear, while the remaining acoustic pattern (the base, which by itself is perceived as a syllable) is presented to the contralateral ear. In this situation, most (but not all) listeners simultaneously experience two distinct percepts: One is the original syllable that would result if the base and the transition were electronically fused; the other is the "chirp" sound produced by the frequency modulation of the isolated transition. Interestingly, if fusion occurs (as it does for the large majority of listeners), the subject hears both the fused speech percept and the nonspeech characteristics of the transition, but not the base by itself. Since the duplicity of perception involves the fused percept and only one aspect of the "prefused" information, the duplex phenomenon does not represent merely a lower vs. higher hierarchical level of information processing, but represents two modes of processing, phonetic and nonphonetic (Liberman, 1982).

In addition to being an interesting experimental demonstration of the two perceptual modes, the duplex phenomenon provides an excellent opportunity to investigate at what level of information processing they become distinct and to understand better the difference between speech and nonspeech perception. An advantage in studying this phenomenon is that the two percepts arise from one and the same physical stimulus, so that all that needs to be manipulated are the instructions to the listener. It has been shown, for example, that presence vs. absence of a preceding interval of silence affects discriminability of formant transitions when they support perception of the consonants "t" and "p" following "s," but has no effect on the discriminability of the same transitions as nonspeech chirps (Liberman, Isenberg, & Rakerd, 1981). This was taken as evidence that the importance of silence in the perception of stop consonants is related to "specifically phonetic (as distinguished from general auditory) processes, and that the effect of silence in such cases is an instance of perception in a distinctively phonetic mode" (Liberman et al., 1981, p. 142). Further studies have shown that, on the speech side of the duplex percept, transitions are perceived categorically, whereas, in contrast, the same transitions heard as "chirps" are discriminated continuously according to onset frequency. Moreover, preposed syllables affect the perception of the transitions when they support the perception of stop consonants on the speech side of the duplex, but not their categorization as nonspeech chirps (Mann & Liberman, 1983).

Studies of duplex perception conducted so far have been aimed primarily at supporting the distinction between phonetic and nonphonetic modes of perception. This aim has been accomplished successfully insofar as it has been possible to manipulate variables that influenced the phonetic side of the duplex percept, but had no effect on the nonphonetic perception of the isolated transitions. Yet, if the nonphonetic mode is truly independent, one should likewise be able to manipulate variables that will selectively influence the perception of the transitions as "chirps," while leaving the speech percept unaffected. A careful investigation of variables that selectively affect each side of the duplex percept is necessary if we are to accept the independence of the two perceptual modes.

To this end, the present study investigated the relation between the phonetic and nonphonetic modes of perception by manipulating factors that have selective effects on one or the other side of the duplex percept. Four experiments are reported. The first examined the effect of the stimulus onset asynchrony (SOA) between the base and the transition, and the effect of white noise auditory backward masking of the transition on labeling each component of the duplex percept. The second experiment examined the effect of attenuating the amplitude of the isolated second formant (F2) transition on labeling each component of the duplex percept. The third experiment presented the white noise mask of Experiment 1 on the same channel as the base and investigated the effects of this masking condition on speech and nonspeech perception of the transition. The fourth and final experiment examined separately the effects of backward masking of formant transitions by white noise and by different formant transitions on the labeling and discrimination of each component of the duplex percept.

Experiment 1

This experiment investigated the effects of increasing SOA on each component of the duplex percept. Specifically, we reasoned that categorization of chirps might be facilitated by an increase in SOA, whereas speech perception is penalized. Cutting (1976) has already reported that when the transition and the base of a CV syllable, [ba], [da], or [ga], are dichotically presented, SOA has a destructive effect on the fusion of the two stimuli. In contrast to this general effect, it should be mentioned that in Cutting's (1976) study the subjects were sometimes able to fuse the transitions and the base even at large SOAs of 80 ms and more as suggested by their above-chance correct labeling of the syllables as [ba], [da], or [ga]. However, one problem with this interpretation of their responses is that the unfused base is frequently perceived as [da], and Cutting's inclusion of this category as a correct response confounded possible nonfusion with successful fusion. We attempted, therefore, as part of the present experiment, to replicate Cutting's results, but to circumvent the problem posed by the ambiguity of [da] percepts in his experiment.

Another goal of the first experiment was to investigate the effect that monotic white noise backward masking of the isolated transitions might have on the speech and the nonspeech aspects of the duplex percept. The assumption underlying the use of backward masking to study perceptual processing is that when the onset of the mask is delayed relative to onset of the target, processing of the target occurs during the delay, but is interrupted by the arrival of the mask (Turvey, 1973). If speech and nonspeech perception represent distinct modes, we might be able to mask selectively one or the other of the

two perceptual aspects of the F2 transition. We would then be able to investigate the level in the auditory system at which the two modes become separated.

The decision to employ a white noise mask was based on several considerations. We wanted a mask that would be most likely to have a selective effect on the nonspeech aspect of the transition. Massaro (1970) has shown that a tone is masked efficiently by a different tone even when the similarity between the pitch of the mask and the target tones is varied over a considerable range. On the other hand, white noise does not mask categorization of one or two tones when the intensity of the mask does not exceed that of the test tones (Kallman & Brown 1983), although it does effectively mask detection of tones (Elliot, 1967). No similar results are available for the masking of CV syllables. However, we assumed that the general unpatterned nature of the white noise mask would minimize the possibility that it interferes with extraction of the phonetic information embedded in the transition, while still interfering with perception of its nonspeech aspects.

The first experiment, therefore, had two goals: (a) to investigate the effects of increasing SOA between the F2 transition and the base, and (b) to investigate the effects of a white noise backward mask presented in the same channel as the transition on subjects' ability to label the speech and nonspeech aspects of the duplex percept.

Method

Stimuli. The stimuli used to create the duplex percepts are schematically represented in Figure 1. They were adapted from two-formant synthetic approximations to the syllables [ba] and [ga], as produced on the parallel resonance synthesizer at Haskins Laboratories. The pattern illustrated in the left panel of Figure 1, which we refer to as the "base," is the constant portion of the two syllables. Its duration is 300 ms, with a 25 ms amplitude ramp at onset, a 100 ms amplitude ramp at offset, and a fundamental frequency that falls linearly from 114 to 79 Hz. The first formant begins at 100 Hz, and during the first 50 ms it increases linearly to achieve a steady-state frequency of 765 Hz. The remaining two patterns, illustrated in the right-hand panel of Figure 1, are the F2 transitions appropiate for [ba] and [ga]. Each was synthesized separately from the base, and is 50 ms in duration. Their common offset frequency is the steady-state frequency of the F2 of the base (1230 Hz), and amplitude contour and fundamental frequency are identical to that of the first 50 ms of the base. The [ba] transition starts at 924 Hz, has a rising frequency contour, and if electronically fused with the base, supports perception of [ba]. The [ga] transition starts at 2298 Hz, has a falling frequency contour, and if electronically fused with the base, supports perception of [ga]. The base alone tends to be perceived as a poor quality [da].

An additional stimulus was created for the purpose of backward masking the perception of the F2 transitions. It consisted of 15 ms of white noise at intensity 1.8 dB above the maximal intensity of the transitions.

Test tapes. The base, F2 transitions, and white noise mask were digitized at 10 kHz, and subsequently recorded onto magnetic tape. Five stimulus series were created: Three practice series to acquaint subjects with the duplex percept, and two test series to assess the influence of temporal asyn-

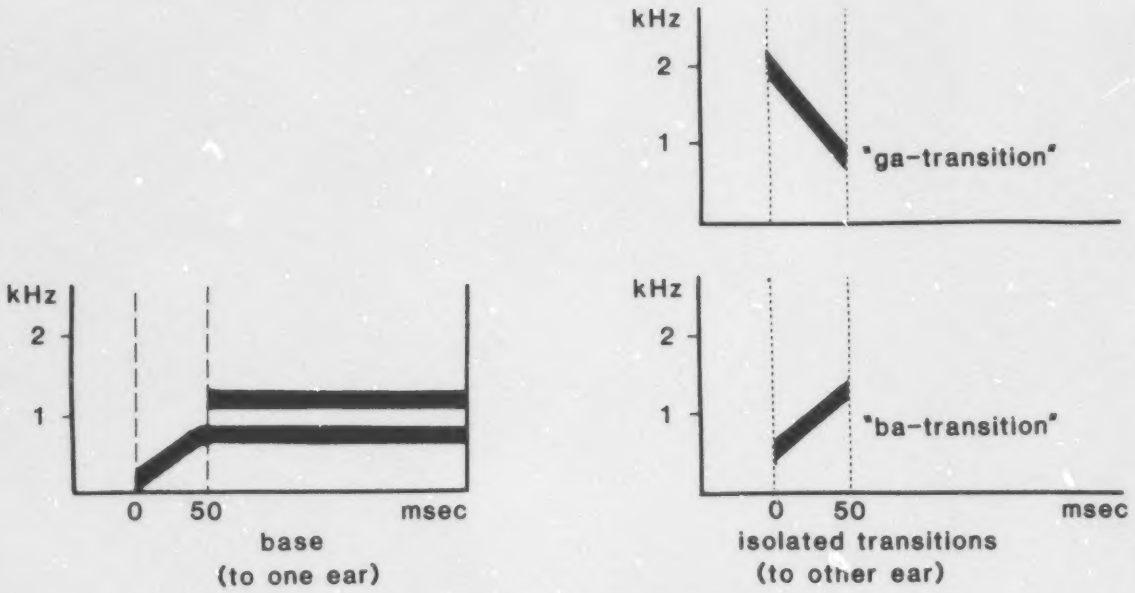


Figure 1. Schematic representation of the patterns used to produce the duplex percepts.

chrony and of the white noise mask on the speech and nonspeech components of the duplex percept.

In the first practice series, designed to familiarize subjects with the speech component of the duplex percept, the base was electronically fused with each of the F2 transitions so as to form two syllables, [ba] and [ga]. These were recorded five times each, and then ten times in alternation. In the second practice series, designed to familiarize the subject with the chirp component of the duplex percept, the isolated [ba] and [ga] transitions were recorded five times each and then ten times in alternation. The third and final practice series was designed to familiarize the subjects with duplex percepts. In it, the base and the F2 transitions were recorded onto separate channels of the tape so as to permit dichotic presentation of each transition in synchrony with the base. These two duplex stimuli were also recorded five times, then alternated ten times.

The two test series included only dichotic stimuli. As in the third practice series, the base and the transitions were recorded onto separate channels, but the synchrony of base and transition was systematically manipulated. In the first test series, the [ba] and [ga] transitions each preceded the base eight times at eight different SOAs: 0, 20, 40, 60, 70, 80, 90, and 100 ms. This yielded a total of 128 stimuli that were recorded in randomized sequence with interstimulus intervals of 2.5 sec, and longer pauses between blocks of 16 stimuli. In the second test series, synchrony of base and transition was again manipulated, but each transition was also immediately

followed by the white noise masking stimulus. Each transition preceded the base eight times at three different intervals: 0, 20, and 40 ms. This yielded a total of 48 stimuli, recorded in randomized series with the same interstimulus and interblock intervals as in the first test series.

Procedure

Subjects listened to stimuli over TDH-39 headphones in a quiet room. They were naive as to the nature of the experiment, being told only that its purpose was to examine whether perception of speech and nonspeech could be altered by certain distractor sounds. They were further advised to attend to the percept designated by the experimenter and to ignore all else. The experiment began with a pretest that involved presentation of the three practice series: the electronically fused syllables, the isolated transitions, and the dichotic stimuli. Participation in the experiment proper required that a subject be able to distinguish accurately the speech and nonspeech percepts presented in the first two practice series, and to label accurately the two components of each duplex percept in the third practice series. In this manner we insured that each subject who continued in the experiment was able to perceive the distinction between the [ba] and [ga] speech percepts and the distinction between the rising and the falling chirps, which were the nonspeech percepts in the duplex listening condition.

Those subjects who met the pretest requirements went on to participate in two experimental sessions, with order counterbalanced across subjects. In one session, their task was to label the speech percepts as containing "b" or "g," while ignoring the nonspeech percepts. That session began with presentation of the first (electronically fused syllables) and third (duplex percepts) practice series, followed by the two test series in counterbalanced order. In the other session, the task was to label the nonspeech "chirp" percept as rising or falling, while ignoring the speech percepts. In this case presentation of the second (isolated transitions) and the third (duplex percepts) practice series preceded the two test series.

Subjects. The subjects who met the requirements of the experiment were ten young women who attended Bryn Mawr College. Two additional subjects were screened but not included in the subject population because they failed to distinguish the [ba] and [ga] components of the duplex percept.

Results

The data for the first experiment comprise labeling responses to the two components of the duplex percept: the speech percept of [ba] or [ga] and the nonspeech percept of a rising or falling chirp. In the first test series we had systematically manipulated the synchrony of the constant base and the variable F2 transition heard in the other ear. The percent correct responses for the speech and nonspeech percepts averaged across the ten subjects appear in Figure 2 as a function of SOA.

In general, subjects were more accurate in labeling nonspeech percepts of rising and falling chirps than in labeling speech percepts of [ba] and [ga]. This difference was significant in a repeated measure ANOVA, $F(1,9) = 12.39$, $MSe = 780$, $p < .005$. The systematic increases in SOA also had an effect on response accuracy, $F(7,63) = 3.39$, $MSe = 65$, $p < .004$, but most importantly, manipulations of temporal asynchrony had opposite effects on speech and non-

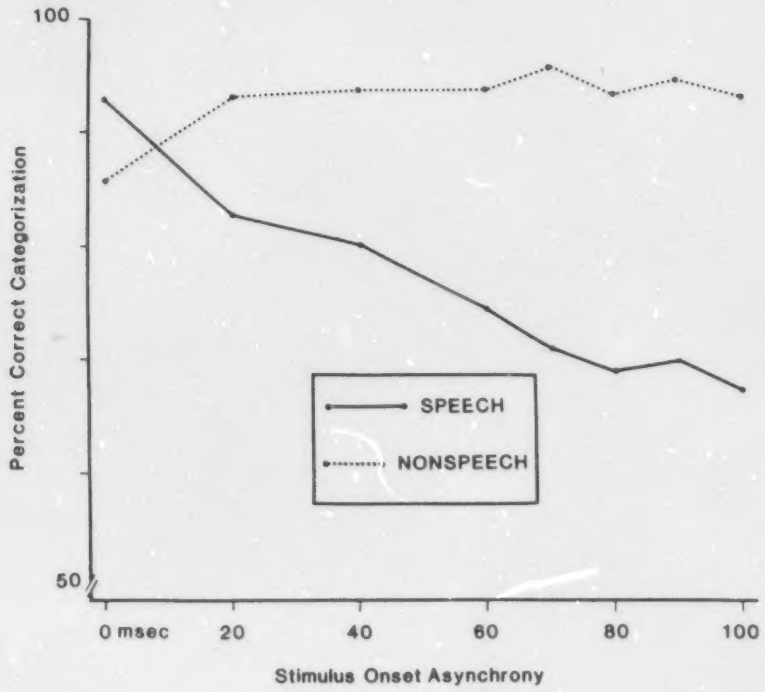


Figure 2. The effects of increasing SOA on labeling syllables and "chirp" categorization.

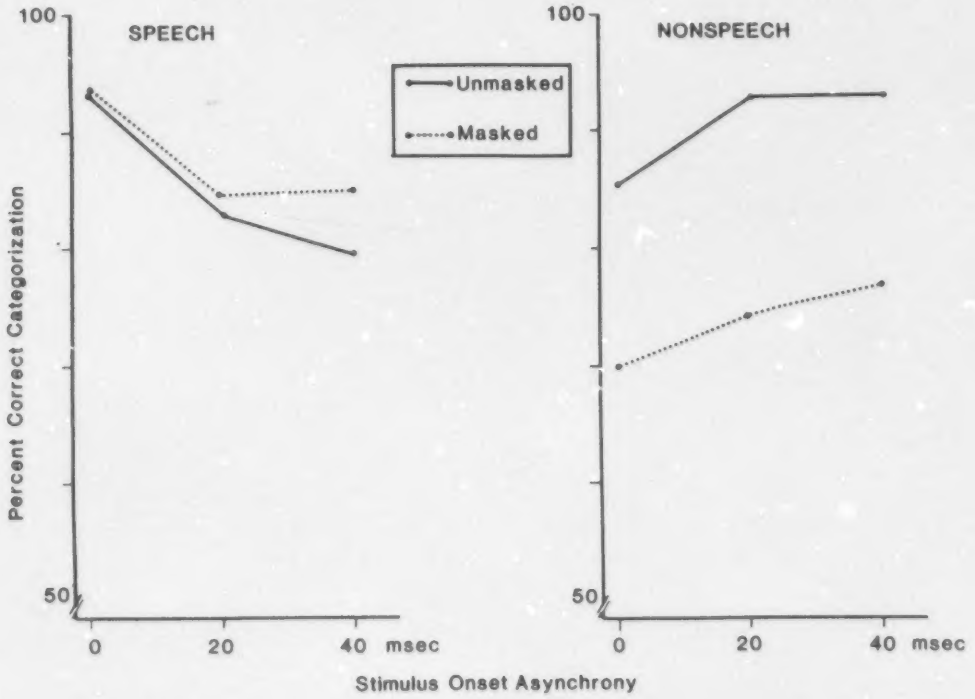


Figure 3. The effects of backward masking the transition by white noise on labeling syllables and "chirp" categorization.

speech percepts: $F(7,63) = 8.75$, MSe = 73, $p < .0001$. Whereas speech perception was best when the base and transitions were in time synchrony, nonspeech perception was better when the transitions preceded the base by 20 ms or more.

We turn now to the results obtained with the second test series, designed to evaluate the effects of backward masking on each component of the duplex percept. The results are presented in Figure 3 where, for convenience, we have also given the results obtained with the comparable unmasked stimuli from the first test series. An analysis of variance computed on the results summarized in Figure 3 revealed that while the white noise stimulus had some effect on performance, $F(1,9) = 7.59$, MSe = 269, $p < .02$, the more interesting result is that the mask penalized perception of the transitions as rising or falling chirps, but had no debilitating effect on the perception of the transitions as cues for distinguishing [ba] and [ga], $F(1,9) = 53.47$, MSe = 72, $p < .0001$. Also, no interaction was found between the SOA and the masking effects. The contrasting influences of backward masking with white noise and manipulation of temporal synchrony on the two components of the duplex percept are to be regarded as the major outcome of this experiment.

Discussion

This experiment has shown that SOA and a white noise mask had selective (but opposed) effects on the speech and nonspeech aspects of the duplex percept. While an increase in the SOA between the base and the transition systematically degraded speech identification when the transition was incorporated into a CV percept, SOAs beyond 20 ms had a facilitatory effect on the labeling of the same transition as a nonspeech chirp. In contrast, a white noise mask that immediately followed the isolated transition had no effect on speech identification, but considerably impaired the discriminability of the chirps.

The explanation of the differential effect of increasing SOA on speech and chirp discrimination is straightforward. SOA has an adverse influence on fusion, and emphasizes the individuality of each channel. Obviously, since chirp identification is based on one channel only, it could not be penalized by this manipulation. Moreover, the higher percent correct categorization of chirps with nonzero SOA might suggest release from a masking effect of the (lower frequency) first formant transition, which might survive in spite of the dichotic presentation. Since identification of [ba] or [ga] is based on fusion of the base and the transition, obviously SOA impaired peception of the syllables. We note, however, that correct labeling of [ba] and [ga] was above chance even when the SOA between the transition and the base was as long as 100 ms. Since the base was identical for both the [ba] and [ga] duplex percepts, any correct identification of the syllable is contingent upon the phonetic information provided by the transition. Nusbaum et al. (1983) claim that listeners may identify the isolated transitions as speech without integrating them with the base. If indeed labeling of syllables at large SOAs was based on phonetic categorization of the transitions, an above-chance asymptote in performance should have emerged. In contrast, a continuous decline of percent correct was found as SOA was increased, supporting the hypothesis that syllable identification in the duplex situation is in fact based on successful dichotic fusion (Repp, in press; Repp, Milburn, & Ashkenas, 1983). We therefore reject Nusbaum et al.'s hypothesis, and assume that some fusion did occur even when the offset of the transition preceded the onset of

the base by 50 ms. Successful fusion implies that, at some point in time, the two stimuli were simultaneously available to the phonetic processor and suggests that the information provided by the leading transitions was somehow stored in memory. The location and form of this storage is not revealed by the present experiment.

The second and, perhaps, more important result is the differential effect of the white noise mask on the speech and nonspeech aspects of the perception of the transitions. The white noise mask impaired categorization of the nonphonetic differences between the two transitions, but the different phonetic percepts supported by them remained unaffected. This result might be explained in two ways. The first is to assume that labeling in the phonetic mode, being both natural and based on highly overlearned categories, require less precise auditory information than does labeling in the nonspeech mode, where the artificiality of the task and the higher amount of uncertainty require more information to be resolved. If this were so, the mask would have had a general effect on the input of the auditory information interfering with a common sensory, precategorical storage mechanism accessed by both the phonetic and nonphonetic processing systems. However, other alternatives should also be considered. A second possible explanation is that there exist different phonetic and nonphonetic information processing mechanisms (Cutting & Pisoni, 1978). If this were so, the unpatterned white noise would have interfered selectively with nonphonetic processing. One way to discriminate between the two explanations is to determine whether other stimulus degradations penalize nonspeech to a greater extent than speech perception. This was tested in Experiment 2.

Experiment 2

One explanation for the differential effect of the white noise mask in Experiment 1 was based on the assumption that speech can be classified into well-known categories, and therefore is more tolerant than nonspeech of the ambiguity induced by the mask in the auditory stimulus. If this is so, other forms of stimulus degradation might have effects on speech and nonspeech perception, similar to those found in Experiment 1. A direct test of this hypothesis was attempted in Experiment 2.

In the first description of the duplex perception phenomenon, Rand (1974) reported that a 30-dB attenuation of the transition relative to the base did not impair correct labeling of the fused speech percepts. With 50-dB attenuation of the transition, labeling performance was still above chance. However, no results were reported regarding the comparable effect of degrading auditory information on the nonspeech aspect of the duplex percept. If correct speech categorization can indeed be based on less auditory information than is required for correct categorization of nonspeech, we should expect that attenuation of the transition will impair speech labeling less than nonspeech categorization. We tested this prediction by gradually attenuating the intensity of the isolated transition for subjects instructed to label the speech or nonspeech aspects of the duplex percept in two separate sessions.

Method

Stimuli. The stimuli employed in Experiment 2 were the base and the [ba] and [ga] transitions of Experiment 1. This time, however, instead of manipulating the temporal asynchrony of base and transition, we kept them in

synchrony as we decreased the relative amplitude of the transition in step sizes of 6 dB from about 80 dB (the amplitude employed in Experiment 1) to 66 dB below that amplitude. The decreases were accomplished on the DDP-224 PCM waveform editing system at Haskins Laboratories.

Test tapes. The three inspection series were as in Experiment 1. In the test series, duplex stimuli were presented, with the base and transition recorded on separate channels in onset synchrony. Each of the two transitions occurred eight times at each of twelve different amplitude levels: equal to the base, -6, -12, -18, -24, -30, -36, -42, -48, -54, -60, and -66 dB. This yielded a total of 192 stimuli, which were recorded in a randomized sequence with interstimulus and interblock intervals as in Experiment 1.

Procedure

The procedure was analogous to that employed in Experiment 1, which had preceded Experiment 2 by several weeks. There were two sessions, with order counterbalanced across subjects. In one session, the task was to label the speech percepts as [ba], [ga], or [da]. The response category of [da] was included because [da] is the response most often assigned to the base in isolation, but no [da] responses were given by our subjets, perhaps owing to their previous experience in Experiment 1. Presentation of the first and third inspection series (electronically fused [ba] and [ga] syllables and duplex stimuli) was followed by presentation of the test series. In the other session, the task was to categorize the percepts as rising or falling in pitch, or to respond with "0" if no nonspeech percept was heard. Presentation of the second (isolated transitions) and third practice series was followed by presentation of the test series.

Subjects. The subjects were the same ten young women who participated in Experiment 1.

Results

The pattern of results, averaged across subjects, is graphically summarized in Figure 4, where the solid line represents the accuracy of responses when subjects were asked to label their speech percepts, the dashed line represents the accuracy of response when subjects labeled their nonspeech percepts of rising and falling chirps, and the dotted line represents the percentage of the trials on which subjects did not hear any chirp at all. In general, the accuracy of speech perception was superior to the accuracy of nonspeech perception, $F(1,9) = 10.85$, MSe = 840, $p < .009$, and the systematic decrease in transition amplitude had a penalizing effect on both, $F(11,100) = 23.23$, MSe = 130, $p < .0001$. Yet, amplitude decreases had a significantly greater effect on nonspeech perception than they had on speech perception, $F(11,99) = 3.11$, MSe = 157, $p < .001$. For example, at an amplitude decrease of 42 dB, subjects reported hearing no chirps at all in 50% of the trials, and their categorization of those chirps that were heard was only 75% correct. Yet, at this same amplitude decrease, speech labeling was 95% accurate. At -48 dB, subjects were reporting a nonspeech percept in only 20% of the trials and their nonspeech labeling was at chance, but their speech labeling was correct in 89% of the trials. Indeed, not until an amplitude decrease of 66 dB did speech perception approach a 50% level of accuracy. This, then, we regard as the major outcome of the third experiment: At certain amplitude levels where nonspeech percepts of the F2 transitions often go

undetected and, even when detected, are categorized at chance levels, speech perception conveyed the phonetic information in the same formant transitions quite accurately.

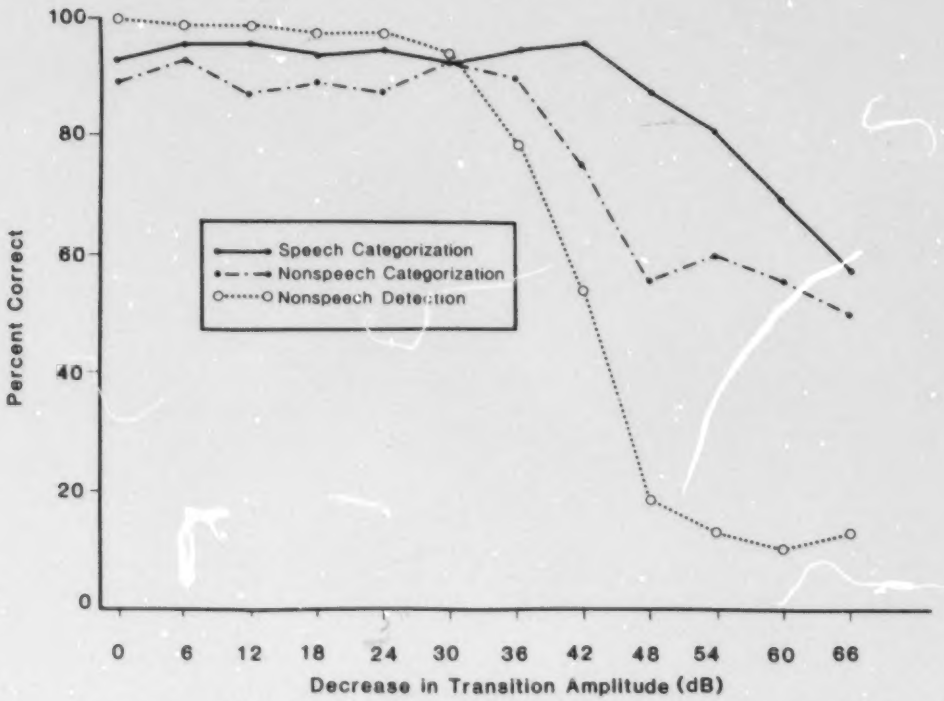


Figure 4. The effects of a decrease in the amplitude of the isolated transition on labeling syllables, "chirp" detection, and "chirp" categorization.

Discussion

The results of Experiment 2 replicate those of Mann (1974) and provide strong support for the hypothesis that categorization of speech may be based on less auditory information than is required for nonspeech categorization. The different sensitivity of the speech and nonspeech perception mechanisms to the decrease in the intensity of the transitions also suggests that they operate at different levels of information processing. As Cutting (1976) pointed out, sensitivity to energy level is indicative of lower-level processing in audition, as Turvey (1973) suggested for visual perception. The smaller effect of the decrease in the transition's energy on speech than on nonspeech perception indicates that speech labeling is less sensitive to changes in energy levels of the stimulus and, therefore, is based on higher-level perceptual processes than is nonspeech categorization.

Experiment 3

In Experiment 1 the mask was presented in the same channel as the transition. The purpose of the third experiment was to examine the effects of a white noise mask, presented simultaneously with the transition but in the same

channel as the base, on speech and nonspeech perception of the dichotic stimulus.

Massaro (1970) reported that contralateral backward masking of tonal targets by tonal masks interferes with pitch perception as much as binaural backward masking. Dichotic backward masking effects have also been found with more complex stimuli, such as CV syllables (Darwin, 1971; Studdert-Kennedy, Shankweiler & Schulman, 1970). In contrast, when the contralateral mask leads the target stimulus, recognition of CV syllables is less impaired (Darwin, 1971; Studdert-Kennedy et al., 1970), and pitch recognition of tones is not affected at all (Massaro, 1970). On the basis of these results and of binaural masking effects, it has been suggested that an auditory input produces a preperceptual auditory image that represents the information in the stimulus and is located centrally. The recognition process entails a readout of the information from this preperceptual auditory image (Massaro, 1972b), and this process is penalized when a lagging stimulus (the mask) occurs before completion of the readout of the necessary information. This hypothesis explains the difference between forward and backward masking effects.

We may assume that two similar intense stimuli simultaneously presented to the two ears will be simultaneously present in the centrally located preperceptual storage. In this case, since the white noise and the isolated chirp are both nonspeech stimuli, they might fuse but should not provide a duplex percept. Since the same "constant" (the white noise) is added to each of the two different chirps, we may expect that the combined noise + chirp percepts would be discriminable and therefore nonspeech categorization will not be affected by the contralateral simultaneous white noise mask. A similar prediction was made about masking effects on the speech aspect of the duplex percept. If indeed speech perception involves readout from the preperceptual image, the base, the transition, and the white noise would all interact. The results of Experiment 1 and 2 suggest, however, that speech information can be extracted quite accurately from a "noisy" image, and therefore the mask should not interfere with syllable labeling.

Method

Stimuli. The stimuli were 30 ms F2 transitions separated from two-formant aproximations of [ba] and [ga], and the base, all newly synthesized on the Haskins software parallel resonance synthesizer. The base was 180 ms in duration, had a 25 ms amplitude ramp at onset, a 100 ms amplitude ramp at offset, and a fundamental frequency that fell linearly from 114 Hz to 81 Hz. The first formant increased linearly from 250 Hz to a steady state of 765 Hz during the first 30 ms. The steady state of the base's F2 was at 1230 Hz and began 30 ms after the onset of the first formant. The [ga] transition fell linearly from 2200 to 1230 Hz, and the [ba] transition rose linearly from 1000 to 1230 Hz. The fundamental frequency of the transitions was identical to that in the first 30 ms of the base.

For duplex presentation in the control condition, a [ba] or a [ga] transition was recorded onto one channel of a magnetic tape and the base was recorded onto the other channel with 30 ms SOA, following the transition. In the masking condition, a 30 ms segment of white noise immediately preceded the base (i.e., the noise was an ipsilateral forward mask with respect to the base, but a simultaneous contralateral mask with respect to the chirp).

Test tapes. A screening tape and a test tape were prepared. The screening tape contained five series of stimuli, starting with the full [ba] and [ga] syllables repeated ten times, followed by ten alternations between the two syllables. In the second series the F2 [ba] and [ga] transitions occurred in a sequence identical to that for the intact syllables. Following the isolated transitions, the duplex percepts were introduced in the following order: First, the base was paired with the [ba] transition and recorded dichotically five times. Next, the base was paired with the [ga] transition and recorded dichotically five times. These two blocks were repeated once, followed by ten alternations between the [ba] and the [ga] duplex stimuli. The final two series on the screening tape were two different randomizations of 30 [ga] and 30 [ba] duplex stimuli. The interstimulus interval (ISI) throughout the tape was 2.5 seconds.

The test tape comprised four different randomizations of 20 [ba] and 20 [ga] duplex stimuli. Two series constituted the control condition, and two the masking condition. The ISI was 2.5 seconds throughout the tape.

Subjects. The subjects were seven out of 12 female subjects screened for participation in Experiment 4. Their only experience with listening to synthetic speech was in Experiment 4 (which was run first), and they were chosen only for reasons of availability. We shall describe here the pretesting procedure, by which the subjects were screened for both experiments.

Twenty-six volunteers were pretested in groups of two to four, in two sessions separated by at least 48 hours. In the "speech" session, subjects were first presented binaurally with the series of [ba] and [ga] syllables, and asked to label them with no restriction whatsoever. All subjects reported that they perceived speech, and used the labels 'ba,' 'da,' 'ga,' and 'ya' to describe their percepts. The blocked duplex stimuli were then presented, followed by the first list of 60 randomized stimuli. The subjects were instructed to label each of the 60 stimuli using their own notation. In the "nonspeech" session, the "chirps" were presented first and the subjects were required to describe the two different percepts. None of the subjects perceived the chirps as speech. The "/" notation was then proposed for the rising chirp ([ba] transition), and the "\" notation for the falling chirp ([ga] transition). The blocked duplex stimuli were then presented again, and the subjects were instructed to label the chirps this time, and to ignore the speech channel. Finally, the second series of 60 randomized duplex stimuli was given, and the subjects again labeled the chirps. Only subjects who were correct on at least 50 out of the 60 trials in both the speech and the nonspeech pretest sessions were included in the experiment proper.

Procedure. Since most stimuli were labeled by all subjects as [ba] or [ga], the speech responses during testing were restricted to those two categories. Subjects were tested in two sessions, separated by at least 48 hours. In one session they labeled the speech percepts, and in the other they categorized the nonspeech percepts as rising or falling chirps. The order of the sessions was counterbalanced among subjects. Each session began with a review, using the screening sequences of syllables and duplex stimuli for the speech sessions, and isolated transitions and duplex stimuli for the nonspeech sessions. Presentation of one randomization for the control and one for the masking condition followed, with the order counterbalanced.

Results

As can be seen in Figure 5, in the control condition speech was correctly labeled 75% of the time, and nonspeech categorization was correct on 76% of the trials. In the masked condition, however, speech perception was significantly impaired, being labeled correctly on only 62% of the trials, whereas correct categorization of nonspeech remained nearly constant at 74%. This condition by perceptual mode interaction was supported by an ANOVA with repeated measures, $F(1,6) = 12.4$, $MSe = 16.2$, $p < .01$.

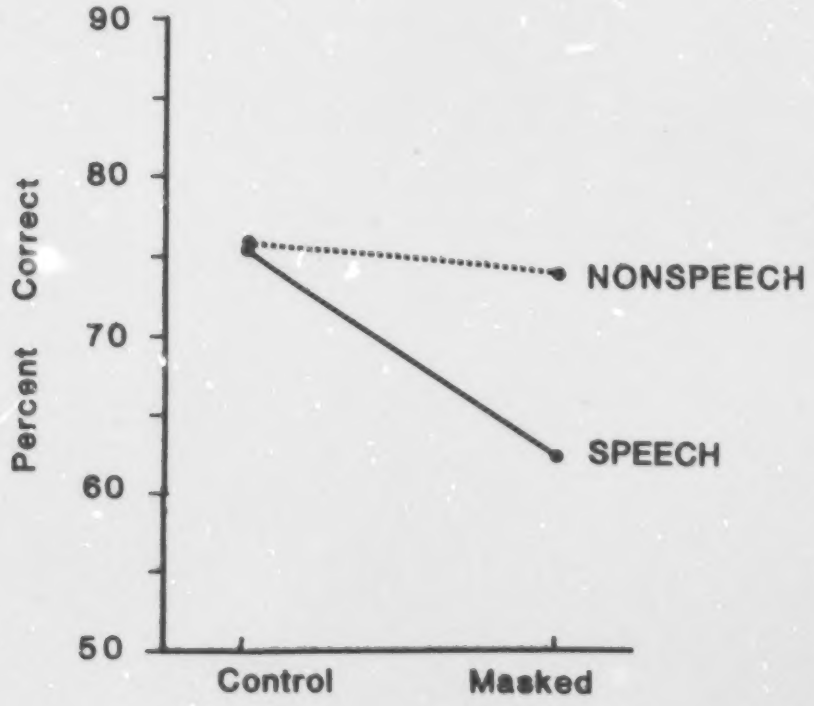


Figure 5. The effects of forward masking the "base" with white noise on labeling syllables and "chirp" categorization.

Discussion

The speech labeling performance in the control condition replicated the results of Experiment 1 at 30 ms SOA, while categorization of unmasked nonspeech was somewhat poorer than expected. This, however, had the advantage that both modes of perception in the control condition were comparable in level of accuracy. When a white noise mask was presented in the speech channel preceding the base, nonspeech perception was not affected while speech labeling was significantly reduced. Thus, simultaneous dichotic masking of the transitions was not effective for nonspeech percepts. Comparing these results with the damaging effect of the monotic backward masking obtained in Experiment 1, and with the effective dichotic backward masking obtained by others, we assume that in the dichotic simultaneous presentation, the transition and the white noise were integrated into one preperceptual image as predicted by Massaro's model, and that this image contained sufficient information to support identification of rising vs. falling chirps.

The masking effect of white noise on speech was not predicted. In line with the results of Experiments 1 and 2, which suggested that speech perception is relatively tolerant of stimulus degradation, we expected no white noise masking of speech, regardless of the channel. Given the reduced forward masking effects reported previously, and the redundant nature of the information in the base, it seems unlikely that the decrease of speech perception performance in this experiment was caused by proactive effects of the white noise mask on the base. However, forward masking effects of the white noise on the F1 transition in the base that might be critical to perception of any stop consonant cannot be excluded. It is also possible that the fused image of the white noise and the transition that was available in the preperceptual storage when the base arrived, might have been qualitatively different from the image of the original transitions and, therefore, less likely to fuse and to provide phonetic infomation to the base. In any event, addition of the white noise altered the recovery of phonetic information in (F1 or F2) transitions but did not penalize recovery of nonspeech information. We cannot, therefore, sustain the conclusion of Experiment 2 that the speech perception system, in general, does not require as precise auditory information as does the nonspeech system. Rather, we should assume that it has different requirements and different sensitivities, and therefore comprises a different mode of information processing. This assumption was further investigated in Experiment 4.

Experiment 4

The fourth experiment was designed to compare the effects of two different backward masks on speech and nonspeech perception: a white noise mask, similar to that used in Experiment 3, and an F2 transition derived from a synthetic approximation to the syllable [da]. The comparison provided a test of the two explanations we have offered to account for the selective effect that the monotic white noise backward mask had on the nonspeech aspect of the duplex percept (Experiments 1 and 2). The first assumed that speech perception may require less information than nonspeech. The second assumed different perceptual processes for speech and for nonspeech, which are sensitive to different aspects of the auditory stimulus.

If some aspects of the auditory stimulus are used by the phonetic perception mechanism to support speech identification, while different aspects of the same stimulus are used for nonspeech categorization, backward masks that contain different amounts of phonetic information might affect the two perceptual modes differently. A white noise mask that provides only limited phonetic information should have little effect on speech perception but, as in Experiment 1, should effectively mask the nonspeech percepts. On the other hand, a [da] transition that is a potential cue for a phonetic percept may more effectively mask speech because it provides phonetically patterned information.

A second means by which Experiment 4 provided a test of the above mentioned hypotheses was by examining the influence of the difficulty of the task on backward masking of speech and nonspeech by the different masks. Thus we compared the relative effects of each mask in a labeling task, similar to the task used in Experiments 1 and 2, and in a discrimination task using the AXB paradigm. In the AXB paradigm, no perceptual predefined categories are necessary. Although the labeling of auditory percepts may facilitate their storage in short-term memory, labels are not essential to accurate discrimination per-

formance (i.e., deciding whether X is A or B). Therefore, we assumed that in the discrimination task, speech and nonspeech perception would be similarly tolerant of ambiguous information, and expected that if selective backward masking effects on speech and nonspeech labeling are due to the overlearning of speech categories, they would be less salient in the AXB paradigm. Obviously, since speech perception may be both sensitive to different aspects of the auditory stimulus and require less information due to overlearning, a three-way interaction between the required perception mode, the nature of the mask, and the difficulty of the mask might exist.

Method

Stimuli. The stimuli for this experiment were two-formant approximations of the syllables [ba] and [ga], newly synthesized on the Haskins software parallel resonance synthesizer. The duration of each of the two transitions was 30 ms, and the duration of the base was 175 ms. The fundamental frequency of the base decreased linearly from 114 to 81 Hz. The first formant rose from 250 Hz to a steady state of 765 Hz. The steady state of F2 was at 1230 Hz. The [ga] transition went from 2200 Hz to 1230 Hz, and the [ba] transition went from 1000 Hz to 1230 Hz. The fundamental frequency contour of the transitions matched those of the first 30 ms of the base.

There were two masking conditions, involving a white noise mask and a [da] chirp mask, and a control condition in which no mask was presented. The SOA between the transition and the base was 30 ms for all trials in all conditions; thus the transition offset coincided with the base onset, and the masks were simultaneous with the first formant transition of the base. In the white noise mask condition, a 30 ms segment of white noise immediately followed the [ba] or the [ga] transitions. In the [da] chirp mask condition, the transitions were immediately followed by a 30-ms F2 transition separated from a two-formant synthetic [da], which had the same base as our test stimuli. The [da] transition fell from 1600 to 1230 Hz, and the fundamental frequency and amplitude contour were identical with those of the [ba] and [ga] transitions.

Tapes. A screening tape and two test tapes were prepared. The screening tape contained five series of stimuli as described in Experiment 3.

The labeling test tape comprised two different series of 120 randomized duplex stimuli (60 [ba] and 60 [ga]). In each series there were 40 control trials and 40 trials for each of the two masking conditions. In both the control and masking conditions, the ba-transition was presented in 20 trials and the ga-transition in the other 20 trials. The ISI was 2.5 seconds throughout the tape.

The discrimination test tape included two different randomizations of 60 test trials, each preceded by eight practice trials. Every trial consisted of three stimuli in AXB design, where the second stimulus, X, was identical either to the first stimulus, A, or to the second, B. The A and B stimuli were, respectively, the [ba] and [ga] duplex stimuli used in the labeling test. Five trials for the control condition and for each masking condition were used in each cell of a counterbalanced design, AAB, ABB, BBA, and BAA. Within each trial all three stimuli were taken from the same condition. The interstimulus interval was 0.5 sec and the intertrial interval was 3 sec.

Procedure

The screening procedures used for selecting subjects in this experiment were described in detail in the Method section of Experiment 3.

Following the screening, speech and nonspeech duplex perception were tested in separate sessions at least two days apart. Speech perception was tested first in half of the subjects, while the other half started with nonspeech. Each session started with a "refresher" series of 30-40 stimuli, in which the subjects practiced labeling the syllables or the chirps. Then they were warned about the presence of the masks, and instructed to label the stimuli using only the [ba] and [ga] labels. The subjects were encouraged to guess each time they were not sure about their percept. The labeling test was always given first, followed by the discrimination test. In the discrimination test, the subjects were instructed to determine whether the first or the last stimulus in each set of three was the "odd one out." They were then given eight practice trials (24 stimuli), followed by the test sequence itself.

Subjects. The subjects were 12 paid female students screened out of 26 volunteers. They had little or no previous experience listening to synthetic speech.

Results

It is noteworthy that during the screening tests all subjects heard the synthetic speech sounds as speech, without any prompting by the experimenter. Also, all but two subjects heard the isolated transitions as nonspeech. (These two subjects were not run in the experiment.) All 12 subjects tested in the experiment proper described the discrimination task as being considerably easier than the labeling task.

Figure 6 presents the percentages of correct responses for speech and nonspeech in each task for the control and masking conditions. A three-factor analysis of variance revealed that performance in the speech mode overall was slightly better than in the nonspeech mode, $F(1,11) = 4.82$, $MSe = 283$, $p < .051$, and that performance was better in discrimination than in labeling, $F(1,11) = 8.00$, $MSe = 371$, $p < .017$. More correct responses were given in the control than in the masking conditions, as revealed by a main effect of condition, $F(2,10) = 27.15$, $MSe = 138$, $p < .001$. Although this main effect was largely due to the white noise mask condition, a post-hoc comparison of the control and the [da] chirp mask conditions turned out to be significant as well, $F(1,11) = 6.89$, $MSe = 73$, $p < .024$. Most importantly, speech and nonspeech perception were differently affected by masking as reflected in a significant interaction between the two factors, $F(2,10) = 6.88$, $MSe = 48$, $p < .014$. Post-hoc analyses of this interaction revealed that for both categorization and discrimination, the white noise mask penalized nonspeech perception more than it penalized speech, while compared with the control condition in each mode, overall, the [da] chirp mask impaired the perception of speech more effectively. The three-way interaction was not significant.

Discussion

Better performance was found in the discrimination than in the labeling task for both speech and nonspeech perception, validating our manipulation of task difficulty. However, the effect of the white noise mask on both speech

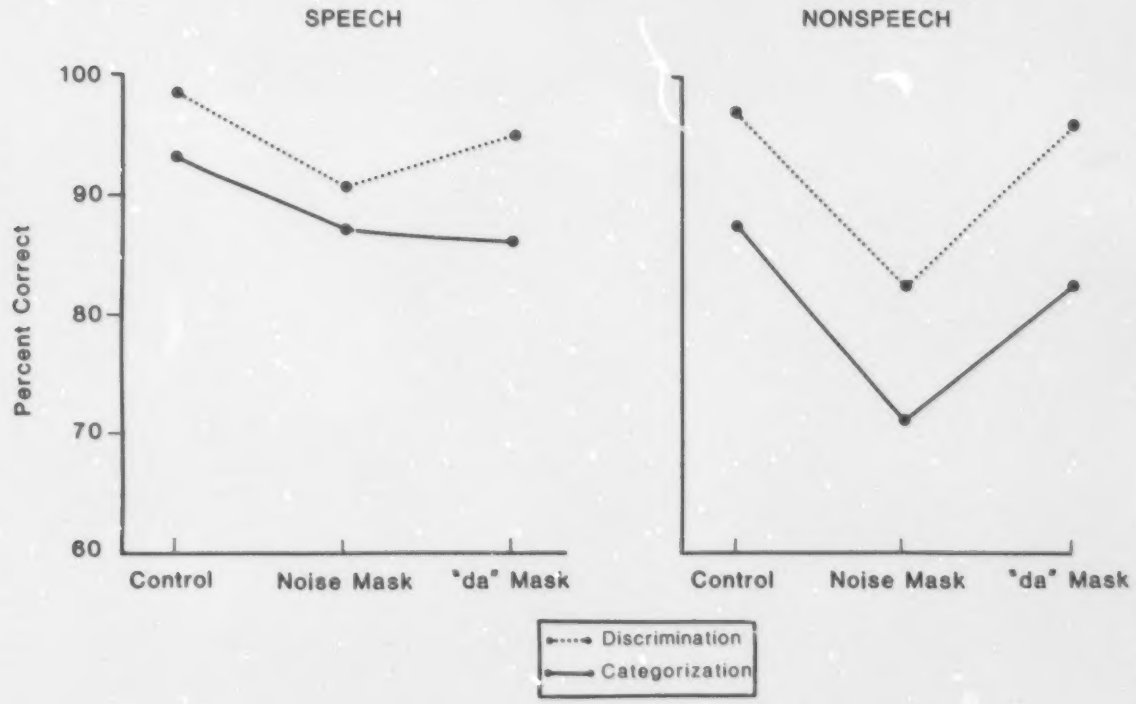


Figure 6. The effects of backward masking the transition with white noise and with a different transition on labeling and discrimination of speech and nonspeech.

and nonspeech perception was identical for both discrimination and labeling of the stimuli. In contrast to Experiment 1, speech perception seemed to be somewhat impaired by the white noise mask, but it was significantly less affected than nonspeech perception. Since the selective effect of the backward white noise mask on nonspeech perception was obtained in both tasks, the results do not support the interpretation that the specific effect of the white noise mask on nonspeech categorization is due to the greater difficulty of nonspeech processing. Rather, we should look for differences in processing phonetic and nonphonetic information that might account for the results.

In contrast to the white noise mask, the effects of the [da] chirp mask are less clear. Since the [da] transition was synchronous with the base, one might expect that this transition would fuse with the base, generating a [da] percept on most trials and reducing any distinction between [ba] and [ga] percepts to chance level. Figure 6 indicates, however, that while the [da] transition reduced correct labeling of speech by 7.7%, correct discrimination was reduced by only 4.3%. It seems that while the white noise mask has a similar effect on discrimination and categorization of both speech and nonspeech, the [da] chirp mask is more efficient in masking speech labeling than speech discrimination. This trend suggests that for speech, the two tasks might have drawn on different strategies and auditory cues.

General Discussion

The present study investigated some differences between speech and nonspeech perceptual modes. Strong additional support was provided for the hypothesis that phonetic and nonphonetic information are processed differently in the auditory system and new information about the nature of this difference was reported.

In Experiment 1 we demonstrated an interesting dissociation between speech and nonspeech perception of a transition presented in a duplex perception context: A burst of white noise following the transition effectively masked the nonspeech aspects of the percept, while leaving intact the phonetic aspects necessary to support perception of a stop consonant. Two possible explanations of this phenomenon were investigated in Experiments 2, 3, and 4. The first, which assumes that speech perception is more tolerant of ambiguous auditory information (since it uses well-learned categories), was apparently supported by the results of Experiment 2. In this experiment we showed that speech perception performance is above chance even when the chirps are no longer detected. However, the results of Experiment 3 revealed that when the white noise mask was presented contralaterally but simultaneously with the transition, speech perception was impaired. We have argued that the white noise could have combined with the second formant to form a new stimulus that was less effective in supporting phonetic perception, or could have masked some critical part of the base, such as the F1 transition. The first interpretation implies that the speech perception system not only is more tolerant of ambiguous auditory input (although it might certainly be so), but also that it is not as sensitive to energy manipulations as it is to certain informational aspects of the stimulus. This hypothesis was tested in Experiment 4, in which the masking effect of the white noise on speech and nonspeech perception was compared with the masking effect of a different transition. Although not entirely conclusive, the results of Experiment 4 supported the following hypothesis: A [da] transition masks labeling of [ba] and [ga] syllables slightly more than white noise, while nonspeech perception is penalized significantly more by white noise than by a transition.

Our finding of successful fusion and accurate CV perception despite SOA between the transition and the base replicates the results of previous studies, and implies the existence of a storage mechanism where the the transition is still available when the base is apprehended. It is with reference to this stage that we will try to explain how the different masks influenced perception of second formant transitions in the speech and nonspeech modes. First, let us consider the effect of backward masking. The white noise backward mask interfered with the readout of auditory, nonphonetic information from the perceptual storage, thereby penalizing identification of the rising and falling chirps. In this case, the effects of the white noise mask were relatively greater than those of the [da] mask because the former was relatively greater in intensity, and we should expect intensity to be of primary relevance in preperceptual, lower-level processing in audition (Cutting, 1976), as well as in vision (Turvey, 1973). The white noise mask, however, did not interfere with phonetic processing of the transition nearly as extensively as it interfered with nonphonetic processing. Perhaps phonetic perception is more resistent to backward masking because the phonetic processor, once triggered by some as yet undefined aspect of an acoustic syllable, is uniquely suited to recovering information in formant transitions that is pertinent to place of articulation. In any case, the particular type of processing involved in

perception in the speech mode would seem to be more vulnerable to the effects of the [da] mask than those of the white noise mask. This could occur because the [da] mask also provides information about the place of articulation, which competes with that provided by the [ba] or [ga] transition. In any event, the masking of phonetic information in the transition is not particularly vulnerable to differences in intensity between the [da] mask and the white noise mask, perhaps because those differences are not relevant to the perception of place of articulation.

Turning now to those results obtained with simultaneous masking, we have found that when a white noise mask is presented that is simultaneous and contralateral to the transition, it has the opposite effect of penalizing phonetic perception, but not nonphonetic perception. Here the greater resistance of nonphonetic perception could be owing to the fact that the noise did not replace the transition in preperceptual acoustic storage, but merely became integrated with it, and hence did not cause a premature cessation of the relevant auditory readout processes. Considering now the penalizing effect of the noise on the speech percept, two explanations have occurred to us. One is that when the transition and the white noise co-occur, they fuse, and the resulting stimulus is less likely to provide phonetic information, which would be fused with that provided by the base. Another is that the white noise obliterates some critical aspect of the base, such as the first-formant transition, which may be a critical cue to stop consonant manner and thus essential to the assigned task of identifying [ba] and [ga].

The results of this study further provide some insight into how, and at what level of information processing, speech is recognized as such, and starts to be processed differentially. Given that once duplex perception is achieved, the base is not heard in and of itself, we are led to entertain the possibility that the storage mechanism is preperceptual, although certain data suggest that it cannot be based on transmission channels (Massaro, 1970). One solution is to adopt Massaro's notion of preperceptual central images and extend it to include auditory cues necessary for phonetic perception. "Phonetic" auditory information stored in the preperceptual storage buffer would not yet be identified as speech. Rather, its ultimate phonetic perception would be contingent upon the activation of a central mechanism. Confining ourselves to the present concern of the distinctions between phonetic and nonphonetic perception of second formant transitions, we would conclude by suggesting, in line with previous processing theories (cf. Turvey, 1973), that speech perception involves activation of a central mechanism, while nonspeech perception is more dependent on peripheral auditory processes. We suggest, therefore, that perception of speech, as a distinct process, starts when the auditory information reaches the central nervous system and "turns on" a special perceptual mechanism.

References

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. Psychological Review, 83, 114-140.

Cutting, J. E., & Pisoni, D. B. (1978). An information-processing approach to speech perception. In J. F. Kavanagh & W. Strange (Eds.), Speech and language in the laboratory, school, and clinic (pp. 38-72). Cambridge, MA: MIT Press.

Darwin, C. J. (1971). Dichotic backward masking of complex sounds. Quarterly Journal of Experimental Psychology, 23, 386-392.

Elliot, L. L. (1967). Development of auditory narrow-band frequency contours. Journal of the Acoustical Society of America, 42, 143-153.

Kallman, H. J., & Brown, S. C. (1983, November). Backward recognition masking of duration and pitch. Paper presented at the twenty-fourth annual meeting of the Psychonomic Society, San Diego, CA.

Liberman, A. M. (1982). On finding that speech is special. American Psychologist, 37, 148-167.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for phonetic mode. Perception & Psychophysics, 30, 133-143.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology (Vol. 8): Perception. New York: Springer Verlag.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.

Massaro, D. W. (1970). Preperceptual auditory images. Journal of Experimental Psychology, 85, 411-417.

Massaro, D. W. (1972a). Stimulus information vs. processing time in auditory pattern recognition. Peception & Psychophysics, 12, 50-56.

Massaro, D. W. (1972b). Perceptual images, processing time and perceptual units in auditory recognition. Psychological Review, 79, 124-145.

Mattingly, I. G., Liberman, A. M., Syrdal, A. M., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. Cognitive Psychology, 2, 131-157.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983). The role of "chirp" identification in duplex perception. Perception & Psychophysics, 33, 323-332.

Rand, T. C. (1974). Dichotic release from masking for speech. Journal of the Acoustical Society of America, 55, 678-680.

Repp, B. H. (in press). Against a role of "chirp" identification in duplex perception. Perception & Psychophysics.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983). Duplex perception: Confirmation of fusion. Perception & Psychophysics, 33, 333-337.

Studdert-Kennedy, M., Shankweiler, D. P. & Schulman, S. (1970). Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. Journal of the Acoustical Society of America, 48, 599-602.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. Psychological Review, 80, 1-52.

VOWELS IN CONSONANTAL CONTEXT ARE PERCEIVED MORE LINGUISTICALLY THAN ARE ISOLATED VOWELS: EVIDENCE FROM AN INDIVIDUAL DIFFERENCES SCALING STUDY*

Brad Rakerd+

Abstract. The purpose of this investigation was to determine whether the presence of neighboring consonants can exert a contextual influence on vowel perception and, if so, to characterize the influence. Two experiments were carried out toward that end. In both, subjects were asked to judge the linguistic similarity relationships that held among a set of American English vowels when those vowels occurred either: (a) in isolation; or (b) in /dVd/ consonantal context. The judgments were made in response to recordings of natural speech in Experiment 1. In Experiment 2, they were made for subjects' memorial images of vowels as elicited by written stimuli. Individual differences scaling of the outcomes of the two experiments provided evidence that supported the following conclusions: (1) consonantal context can significantly influence vowel perception; (2) for the /dVd/ context at least, the nature of the influence is to evoke more linguistic perceptual processing of vowels than occurs when they are presented in isolation; (3) the influence is more likely to be explained in terms of properties of the stimuli presented to perceivers than in terms of any sort of knowledge that perceivers bring to bear in perceptual processing; and (4) three features of linguistic description for vowels--advancement, height, and tenseness--have particular import for vowel perception and for vowel memory.

It has long been recognized that the acoustic correlates of a vowel can vary, sometimes to a substantial degree, depending on the identity of the consonants that precede and/or follow it (e.g., House & Fairbanks, 1953; Lindblom, 1963; Stevens & House, 1963). This variation has come to be understood in terms of the fact that a talker often coarticulates the neighboring segments of an utterance (that is, overlaps their respective productions) such that the acoustic signal is jointly influenced by those segments (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). How, then, do vowel

---

*Also Perception & Psychophysics, 1984, 35, 123-136.

+Also University of Connecticut.

Acknowledgment. This manuscript is based on portions of a doctoral dissertation that was submitted to the University of Connecticut. I am indebted to my major advisor--Robert Verbrugge--and the members of my committee--Alvin Liberman, Michael Turvey, and Benjamin Sachs--for their counsel on every phase of the dissertation project. Much of the research was conducted at Haskins Laboratories and funded by grants awarded to that institution (NICHD grant HD-01994, BRS grant RR05596). I gratefully acknowledge that support, as well as the support of the University of Connecticut Research Foundation.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

perceivers adjust to these acoustic variations? One possibility is that in many or most cases they do not. If the variations are sufficiently minor, a perceiver could simply "ignore" them and achieve an acceptably high level of performance in identifying vowels. Alternatively, the perceiving of vowels might involve certain context-sensitive perceptual strategies, analogous to those that are generally thought to be required when listeners identify the consonants of an utterance (for reviews of the evidence regarding consonantal perception, see, e.g., Liberman, 1982; Liberman et al., 1967; Liberman & Pisoni, 1977).

It becomes important to determine whether or not vowel perceivers are sensitive to consonantal context because (unlike most consonants) vowels can be freed from the influences of their neighboring segments and produced as isolated utterances. A good deal is known about the perception (and production) of these isolated vowels and there is an important issue as to how to generalize from that knowledge to other cases. To the degree that listeners are, in fact, indifferent to the acoustic variations engendered by consonantal context, isolated productions might be taken to be the canonical vowel form, and their acoustic signature to be the one that best exhibits the essential information for vowel perception. However, if the perceiving of vowels does, in general, involve context-sensitive strategies, then the isolated vowel form is but one of many variants, and, arguably, one of the least representative variants because it occurs infrequently in natural speech. An answer to the question of whether isolated vowels are perceived differently than vowels in consonantal context therefore proves to be basic to vowel research.

Previous efforts to answer this question have generally been based on comparisons of the identifiability of vowels in and out of some consonantal frame. Evidence gathered with this method has, in a number of instances, been taken to favor the view that consonantal context can signficantly affect vowel perception by exerting a positive influence on vowel identification (Gottfried & Strange, 1980; Strange, Edman, & Jenkins, 1979; Strange, Verbrugge, Shankweiler, & Edman, 1976). This finding remains a subject of debate, however. It has not been observed in all studies (Macchi, 1980; Pisoni, 1979) and it has been challenged on grounds of being largely an artifact of the method of assessment (Assmann, Nearey, & Hogan, 1982; Diehl, McCusker, & Chapman, 1981; but see Rakerd, Verbrugge, & Shankweiler, in press; Strange & Gottfried, 1980).

The present study complements this work by addressing the question of a consonantal influence on vowel perception with evidence of a different kind than has been offered in the past. To begin with, the data collected here were judgments of vowel similarity rather than absolute identification judgments; hence, they assess the consonantal influence with a new perceptual measure. More importantly, the resulting data were analyzed with an individual differences scaling technique that highlights aspects of the data structure that have not been considered previously. Those aspects are: (1) the dimensions of perception that had some shared significance for the set of subjects as a whole; and (2) the relative salience that those dimensions had for the individual subjects depending on whether they judged vowels in or out of context. It will be seen that both aspects of the scaling solution were informative about the nature of vowel perception.

Experiment 1

The starting point for an individual differences scaling analysis of vowel perception is to collect, for each subject, a matrix of what Shepard (1962a, 1962b) has called "proximities" data. These are data indexing the network of perceptual relationships that hold among a set of vowels. A triadic comparisons procedure was employed to collect those data in the present experiment. That procedure was chosen because it had proven useful in previous vowel research. More than a decade ago, Pols and his associates (Pols, Tromp, & Plomp, 1973; Pols, van der Kamp, & Plomp, 1969) assessed the perceived vowel quality of spectrally-constant speechlike sounds by requiring that subjects compare triplets of stimuli on a trial. Specifically, subjects were required to judge which two members of a triplet sounded most alike to them and which two least alike. They then proceeded to a new triplet and, over trials, judged all possible stimulus combinations. This procedure yielded reliable data that were interpretable, both from a linguistic standpoint and with respect to acoustic properties of the stimuli. Others (Singh & Woods, 1970; Terbeek & Harshman, 1971) have since employed the triadic comparisons method in vowel perception studies and obtained equally satisfactory results. It was used here to compare the perception of isolated vowels with that of vowels in a consonantal context (/dVd/).

Method

Subjects. Twenty-three subjects, randomly selected from a pool of individuals registered with the Haskins Laboratories in New Haven, Connecticut, were paid to participate in Experiment 1. All of them were native speakers of English and none had any history of hearing difficulties. It was ensured that they had no prior knowledge of the purpose of this study or the design of the experiment. Twelve of the subjects were assigned to the isolated-vowels condition of the experiment, eleven to the consonantal-context condition.

Stimuli. The stimuli were natural productions of ten American English vowels: /i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, o, ʊ, u/. A single male talker, who spoke a General American Dialect, recorded these vowels in each of two contexts: (1) in the trisyllable frame /hədVdə/, where the second syllable (/dVd/) was stressed; and (2) in isolation. The /dVd/ consonantal frame was chosen because it imposed certain coarticulatory constraints on the talker. In order to produce initial and final /d/ consonants, the jaw must be closed and the tongue tip sealed against the back of the teeth. Articulation of the syllable vowel, which likewise requires an appropriate parameterization of the tongue and jaw, must therefore be coordinated with that of the consonants. Presumably owing to these coarticulatory constraints, there often is a substantial degree of acoustic modulation associated with /dVd/ syllables. The stressed target syllables were flanked by destressed syllables (/hə/ and /ə/) to ensure that the consonantal-context stimuli would not be meaningful words in English.

While seated in a sound-attenuated room, the talker produced several tokens of each vowel in each context. These productions were tape recorded, low-pass filtered at 5 kHz, digitized at a sampling rate of 10 kHz, and stored in separate computer files. Two of the tokens of each vowel were used in the experiment. In all cases, these were the first two tokens produced unless some sort of articulatory anomaly such as vocal fry or "breathiness" was audi-

ble. When an anomaly was heard in one of the first two tokens, it was replaced by the third, and if that was anomalous by the fourth and so on. Acoustic analyses revealed that the stimuli selected by this procedure were acoustically "normal," in that their spectral and temporal characteristics were such as might be expected on the basis of data reported by previous investigators (e.g., Lehiste & Peterson, 1961; Peterson & Barney, 1952).

Procedure

Instructions. At the outset, it was explained to subjects that the task would be to compare their perceptions of several different vowel sounds. It was also explained that they were to base the comparison on linguistic aspects of those sounds. The individual subjects were left to define their own criteria for the linguistic aspects. They were, however, given the following example as an aid:

> If a child and an adult were both to say the vowel /i/ or the word /did/, you would surely hear some differences between the vowel sounds. The child's vowels would doubtless be softer, higher in pitch, and so on. On the other hand, the /i/ vowels produced by the child and the adult would also have something in common, a quality or qualities that distinguished them from other vowels like the /ɛ/ in /dɛd/ or the /ɪ/ in /dɪd/. These are the qualities that you should attend to in this experiment.

Triadic comparisons. As indicated earlier, the specific task set for subjects was that of comparing triads of stimuli. On each experimental trial, three vowels were randomly selected for presentation from the set of ten alternatives with the constraint that the particular triad chosen had not occurred on any previous trial. A subject was allowed to listen to these three vowels in any order and any number of times with the goal of reporting which two of the three sounded most alike and which two least alike. Over the course of the experiment, listeners judged all possible triadic combinations of the ten vowel alternatives (120 possibilities). Note that this meant that every vowel pair was, over trials, judged in relation to every other vowel in the set.

Data were accumulated over trials according to the following scoring procedure: vowel pairs judged most-alike were assigned +1 scores and those judged least-alike -1 scores. In this way, a matrix of data indexing the perceived relationships among the ten vowel alternatives was obtained for each subject. The matrix for a subject who rated vowels in consonantal context is shown in Table 1 for purposes of example.

One of the virtues of the triadic-comparisons procedure was that it placed minimal memorial demands on subjects, since only three stimuli had to be dealt with on each trial and these could be played and replayed in any order as needed. A second virtue was that the self-paced nature of the procedure minimized the time pressure felt by subjects.

Familiarization with the equipment and procedures. A complete testing session took about two hours. Roughly thirty minutes of that time was devoted to familiarizing subjects with the equipment and procedures used to present the stimuli and record the responses.

Table 1

Similarities matrix for a subject who rated vowels in consonantal context

	i	ɪ	ɛ	æ	ʌ	ɑ	ɔ	o	ʊ	u
i										
ɪ	6									
ɛ	-2	2								
æ	-4	-4	5							
ʌ	-2	-4	7	1						
ɑ	-8	-4	1	4	1					
ɔ	-5	-3	0	1	4	5				
o	-6	-5	-2	0	2	-3	3			
ʊ	-5	1	4	-1	5	-2	1	3		
u	-3	-2	2	-1	1	-1	-4	4	8	

The subjects were tested individually. Each was fitted with headphones and seated in front of a computer terminal that was housed in a sound-attenuated room. Three of the keys on the terminal triggered presentation of the appropriate stimuli for each triadic trial. After a key was pressed, the corresponding stimulus was presented through headphones at a comfortable listening level. The most-alike and least-alike response choices were entered via a different set of keys on the terminal. Once these choices had been made, the system advanced to the next trial.

The equipment and testing procedure were demonstrated to subjects over a series of training trials. There were between 15 and 25 such trials depending on the individual. Each training trial comprised a different triadic combination of stimuli sampled from the set of 120 possibilities. For the first few such trials, the experimenter operated the equipment, directing the presentation of stimuli and entering response choices. After that, control was passed to the subjects and they paced themselves. They were invited to ask questions about all aspects of the procedure. The testing session was begun only after subjects had both demonstrated competence in operating the equipment and expressed confidence about understanding the perceptual task.

Analysis of the Data

To allow for a direct comparison between conditions, a single individual differences scaling analysis was carried out on all subject data from the two experimental conditions combined. The fundamental modeling assumption of individual differences scaling is that when judging the same set of stimulus items, all subjects will make reference to the same perceptual dimensions. Subjects may differ from one another in terms of the relative weight (salience) that they attach to those dimensions, but the cannot differ in terms of the identity of the dimensions themselves (Carroll & Chang, 1970; Wish & Carroll, 1974).

Consistent with this assumption, the scaling solution has two components that, together, optimally account for the data structure of the individual subject matrices. The first component, called a group space, is a model of what the subjects have in common. The axes of the group space are the shared perceptual dimensions, and these index a set of appropriately positioned points representing the stimulus items. The second component of the scaling solution is a weight space, which specifies the relative salience that the several dimensions of the group space have for each subject. More formally, a subject's weight for a particular dimension reflects the amount of variance in her/his data that can be accounted for in terms of that dimension. Together, the set of weights index a subject's location in the weight space.

A noteworthy property of individual differences scaling is that it dictates the orientation in which the scaling solution must be interpreted. When attempting to relate the solution to potentially relevant factors, an investigator is obliged to make reference to the shared perceptual dimensions. These dimensions enjoy a priority because they are the ones that account for the greatest percentage of variance in the several subjects' data. The presence of this interpretive restriction sets individual differences scaling apart from other variants of multidimensional scaling[2] and strengthens claims that the dimensions of its scaling solution have some psychological reality (Carroll & Wish, 1974; Kruskal & Wish, 1978; Wish & Carroll, 1974).

It has often been pointed out (e.g., Kruskal & Wish, 1978; Wish & Carroll, 1974) that since multidimensional scaling methods are intended to aid in the description and understanding of data, evaluation of the "correctness" of various scaling decisions must be based on considerations that are substantive as well as statistical. The substantive considerations have to do with such factors as the interpretability and stability of results, the statistical with the goodness-of-fit between model and data. On the basis of these considerations, it was determined that the present data were most appropriately modeled: (a) in three dimensions; and (b) at the ordinal scale of measurement.

Dimensionality of the space. With individual-differences scaling, a commonly used index to goodness-of-fit is the percentage of variance accounted for (VAF) in the several subjects' data (see, for example, Carroll & Chang, 1970).[3] Increasing the number of dimensions will increase the VAF index, since the model has added degrees of freedom with which to fit the data.[4] The gains tend to diminish exponentially, however, and each new increment must be weighed against the substantive considerations mentioned earlier (interpretability and stability). Figure 1 displays the VAF function for the present data when modeled in two to five dimensions. The exponential nature of the function is clear. There is a relatively large increment in VAF for the shift from two to three dimensions, a much smaller one for the shift from three to four dimensions, and a negligible decrement (see footnotes 3 and 4) for the shift from four to five dimensions.

On the basis of these statistical data, at least, it appears that three or perhaps four dimensions would be the appropriate modeling choice. In the former case, 70% of the variance would be accounted for, in the latter, 72%.

The three-dimensional solution was chosen over the four-dimensional for two reasons. First, as will be seen shortly, all three of the dimensions are linguistically meaningful and therefore interpretable, and second, they are stable in that they emerged, as well, from analyses of individual subject data

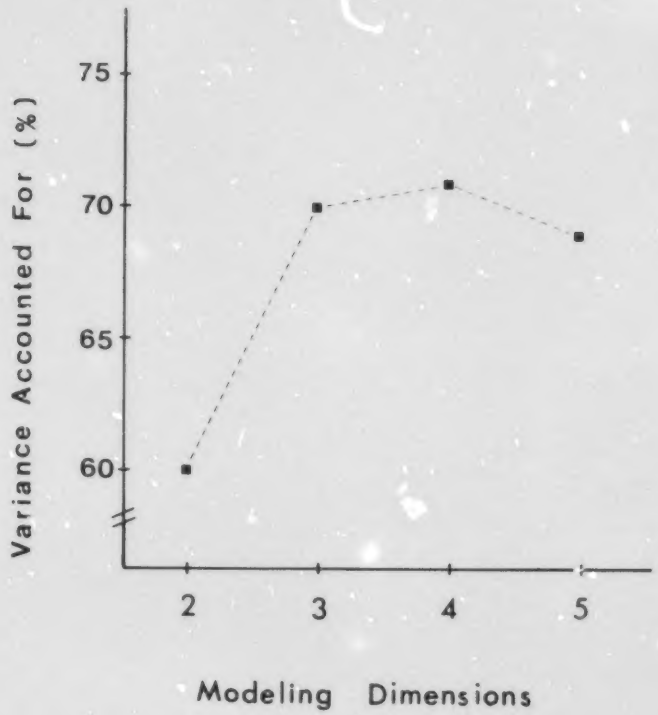


Figure 1. Percentage of variance in the perception data accounted for by modeling in from two to five dimensions.

collected in a memory experiment (Experiment 2) and from separate analyses of the two perceptual conditions of this experiment.

Nonmetric scaling. The decision to model these data at the nonmetric (i.e., ordinal) scale of measurement was based on two considerations. The first was simply that the subjects' task was to make ordinal perceptual judgments. On each trial, it was required that they identify the most-alike and least-alike vowel pairs, but it was not required that they quantify the strengths of those pairwise relationships. It is true that by summing over trials some quantification was arrived at, but it was felt that the most conservative treatment of these data was to model their ranks.

The second reason for operating at the nonmetric scale has to do with the stability of the modeling outcome. The metric/nonmetric distinction made relatively little difference with respect to the present data, but it greatly affected the outcome of modeling the results of a memory experiment (Experiment 2) that was run, in part, to clarify the findings of the present perceptual experiment. This point will be discussed in greater detail later (see the section on Analysis of the Data of Experiment 2). For now, it is enough to note that a nonmetric scaling of these data revealed structure that was stably present for both of the conditions of Experiments 1 and 2.

Results

Group space. The group space for all subjects (isolated-vowels and consonantal-context conditions combined) is shown in Figure 2. Dimension 2 is plotted against dimension 1 in the top half of this figure, dimension 3 against 1 in the bottom half. These dimensions are orthogonal to one another and can be considered independently. To do this, it is useful to "project" the points visually onto each axis and consider their ordering. In this way, it was determined that each of the dimensions of the group space corresponds closely to a traditional feature of linguistic description for vowels. Dimension 1, for instance, corresponds to a feature linguists have variously called advancement, front/back, and grave/acute. After Singh and Woods (1970), the term advancement will be used here. This feature distinguishes vowels such as /i, ɪ, ɛ, æ/ (seen to "project" onto the lower end of dimension 1) from other vowels such as /ʌ, ɑ, ɔ, o, ʊ, u/ (which "project" onto the upper end of the dimension). Dimension 2, in turn, corresponds to what has been called the height or compactness feature (height will be used here), and dimension 3 to tenseness or length (tenseness is the term that will be used).[5] (See, e.g., Hockett, 1958, Jakobson & Waugh, 1979, Ladefoged, 1971, 1975, for comprehensive reviews of the vocabulary of vowel feature description.)

These features repeatedly surface when linguists try to document various aspects of linguistic behavior. To take just one example, speakers of English, though generally unaware of it, observe a grammatical rule for vowel usage that respects the tenseness feature: In English, words can end with "tense" vowels like /i, ɑ, u/ (there are words like "he," "go," and "you"), but they cannot end in "lax" vowels like /ɪ, ɛ, ʊ/ (there are no words ending in the vowel sounds heard in the middle of words like "hit," "bet," and "book"). English speakers must be at least tacitly aware of this rule, since they respect it when creating new words for the language. There are countless other instances of linguistic behavior that is systematically related not only to the tenseness feature but to the advancement and height features as well (see, e.g., Jakobson & Waugh, 1979).

There is, as well, some evidence to support the claim that all of these features play a perceptual role (e.g., Hanson, 1967; Shepard, 1972; Singh & Woods, 1970). The present results are both consistent with such a claim and particularly compelling in this regard given the nature of the scaling analysis that was employed here. Individual differences in the several subjects' data provided information that allowed for a nonarbitrary determination of the dimensions of the group space. Those dimensions shown in Figure 2 are the three that optimally accounted for the variance in the linguistic judgments made by the subjects who participated in Experiment 1. The fact that each of those dimensions, in turn, corresponds closely to a linguistic feature, strongly suggests that those features have some shared perceptual significance among speakers of English (see Rakerd, 1982, for an expanded consideration of this point).

Weight space. The concern in this section will be to look at individual differences in weighting of the dimensions of the group space, and, in particular, at differences between the isolated-vowels and consonantal-context subjects. The weight space for all subjects is shown in Figure 3. Weightings for dimension 2 are plotted against those for dimension 1 at the top of the figure; dimension 3 weightings are plotted against dimension 1 weightings at the bottom. Each O represents an individual from the isolated-vowels condition, each X represents an individual from the consonantal-context condition.

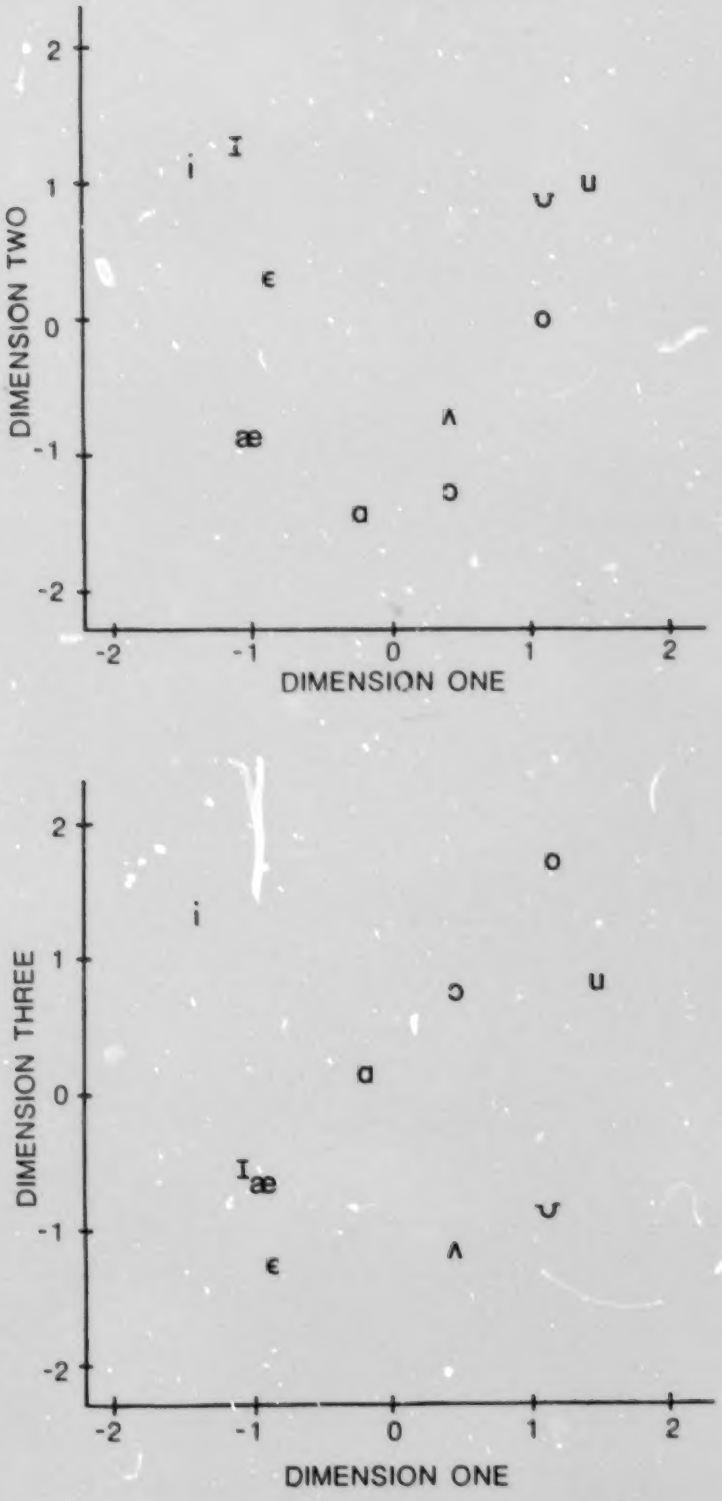


Figure 2. Group space for Experiment 1.

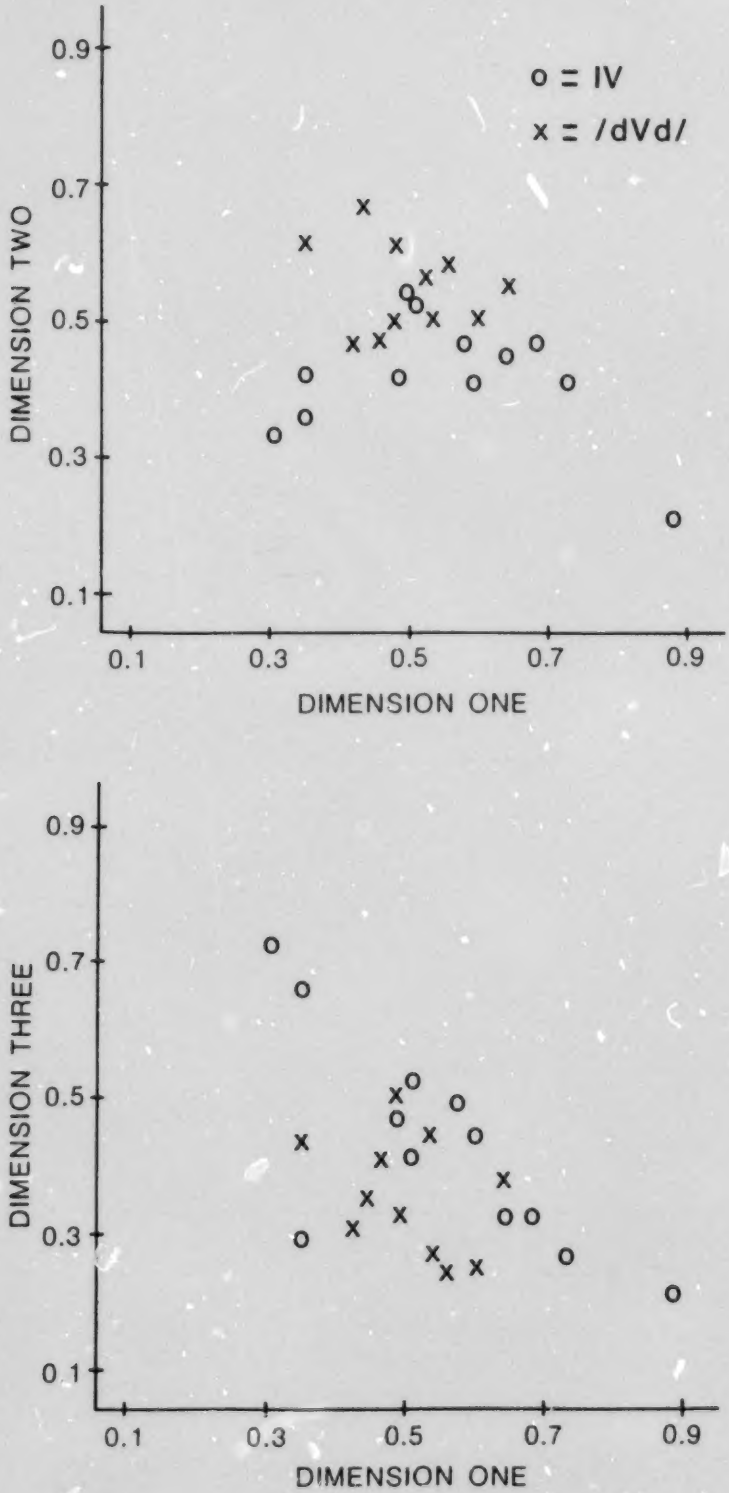


Figure 3. Weight space for Experiment 1.

The first thing to notice in this figure is that the X's are more tightly grouped than the O's in both the dimension 2--by--dimension 1 and dimension 3--by--dimension 1 planes. This indicates that there was less variability among consonantal-context subjects than there was among isolated-vowels subjects. In order to assess the statistical significance of this difference, a variance ratio (Snedecor's F) was calculated for the three-dimensional space as a whole. For each condition the variance in three-space was determined as follows. First, the centroid, or average subject weight was located in the space. Then, the distance from this centroid was calculated for each subject according to the Pythagorean theorem.[6] And finally, the average squared distance was computed as the measure of variance. This measure is strictly analogous to the variance statistic (sigma squared), which is the average squared deviation about the mean for a set of numbers. The difference in variability between the two experimental conditions was in fact significant, $F(11,10) = 3.21$, $p < .05$.

It was observed, then, that in a perceptual task in which listeners were asked to relate a set of vowel sounds on the basis of their linguistic qualities, there was significantly greater agreement among individuals who heard vowels in a consonantal context than there was among those who heard isolated vowels. It can be inferred from this that the subjects employed somewhat different perceptual strategies in the two conditions. This, in turn, suggests the need for caution in generalizing from what is known about the perception of isolated vowels to the perception of vowels in context, a generalization that has often been made in the past (e.g., Chiba & Kajiyama, 1958; Joos, 1948). Also, it would seem to warrant a methodological caveat for those who do vowel research in the future: namely, that they would do well to look at vowels in consonantal context. Given the importance of these implications, it was deemed appropriate to look at the stability of this result, and to ensure that it was not an artifact of the scaling procedure.

Regarding scaling, it is noteworthy that part of the variability in the weight space reflects differences in the goodness-of-fit between the scaling model and the individual data. Subjects whose data were well fit by the model lie further from the origin of the space (in some direction) than do those whose data were poorly fit. It may be that the observed condition difference in variability was, in fact, a difference with respect to goodness-of-fit. One reason for believing this was not the case, however, is that, on average, the subject data in the two conditions were about equally well fit by the model. The average VAF for the isolated-vowels condition was 71%, and that for the consonantal-context condition was 69%, a difference that did not even approach significance.

Whether goodness-of-fit was significantly different for the two groups or not, it was certainly a source of variation that is of limited interest here. Therefore, the data were transformed to "factor out" its influence. A subject's weight on a dimension is the square root of the percentage of variance accounted for by that dimension. The total variance accounted for (VAF) can thus be computed for any individual by squaring the weights and summing over all three dimensions. Between-subject differences in goodness-of-fit can, in turn, be compensated for by normalizing the data with respect to this VAF value. The most straightforward strategy for doing this is to divide a subjects' squared dimension weights by VAF and to take the square roots of the resultant dividends to be the adjusted weights.

These new values index subjects in the weight space shown in Figure 4. It can be seen to reflect statistical compensation for goodness-of-fit differences between subjects, in that the original weights (shown in Figure 3) have been "compressed" along lines extending out from the origin of the space. Despite this compensation, the condition difference in subject variability remains significant, $F(11,10) = 3.18$; $p < .05$.

It proves to be the case, then, that even when individual differences in the goodness-of-fit of the model are factored out, there remains a significant difference between the two experimental conditions with respect to the subject variability in the weight space.[7] This finding clearly supports the view that vowels are perceived significantly differently in consonantal context than out. It also hints at the nature of the difference, at least for the present experiment. The task set for subjects was to relate a number of different vowel sounds on the basis of their linguistic qualities. Subjects who heard vowels in /dVd/ consonantal context exhibited significantly greater agreement as to what those linguistic relations were than did their counterparts who heard isolated vowels. Thus, it can be said that one of the effects of context was to stabilize linguistic judgments across subjects.

It is useful to take note of the nature of the stability; the consonantal-context subjects clustered toward the center of the weight space, which indicates that they attached roughly equal weight to all three linguistically-meaningful dimensions of the group space. Notice that this need not necessarily have been the case. Between-subjects agreement would have been equally high for this condition had the clustering occurred out near one of the "corners" of the weight space, whereupon one or another of the perceptual dimensions could have been said to predominate. It turned out, however, that all three of the dimensions had substantial perceptual import for consonantal-context subjects.

The situation was markedly different for the isolated-vowels subjects. While several members of that group were positioned near the center of the weight space, most of them were in more "extreme" locations. The data for this latter group were largely accounted for in terms of perceptual sensitivity to just one or two of the linguistic dimensions. And it should be noted that the one or two dimensions that predominated were different for different individuals. Thus it can be seen that isolated-vowels subjects were not constrained to perceive the stimuli in terms of the full set of linguistic dimensions. To the contrary, they attended to the dimensions in a piecewise manner, while the consonantal-context subjects integrated the dimensions in a more linguistically-appropriate way.

Summary. The individual differences scaling analysis revealed two ways in which vowels in consonantal context can be said to have been perceived more linguistically than isolated vowels. First of all, judgments about the linguistic qualities of vowel sounds were significantly more stable across subjects when the vowels were in context. And second, three linguistically-meaningful dimensions of vowels were more integrated in perception when vowels were in context.

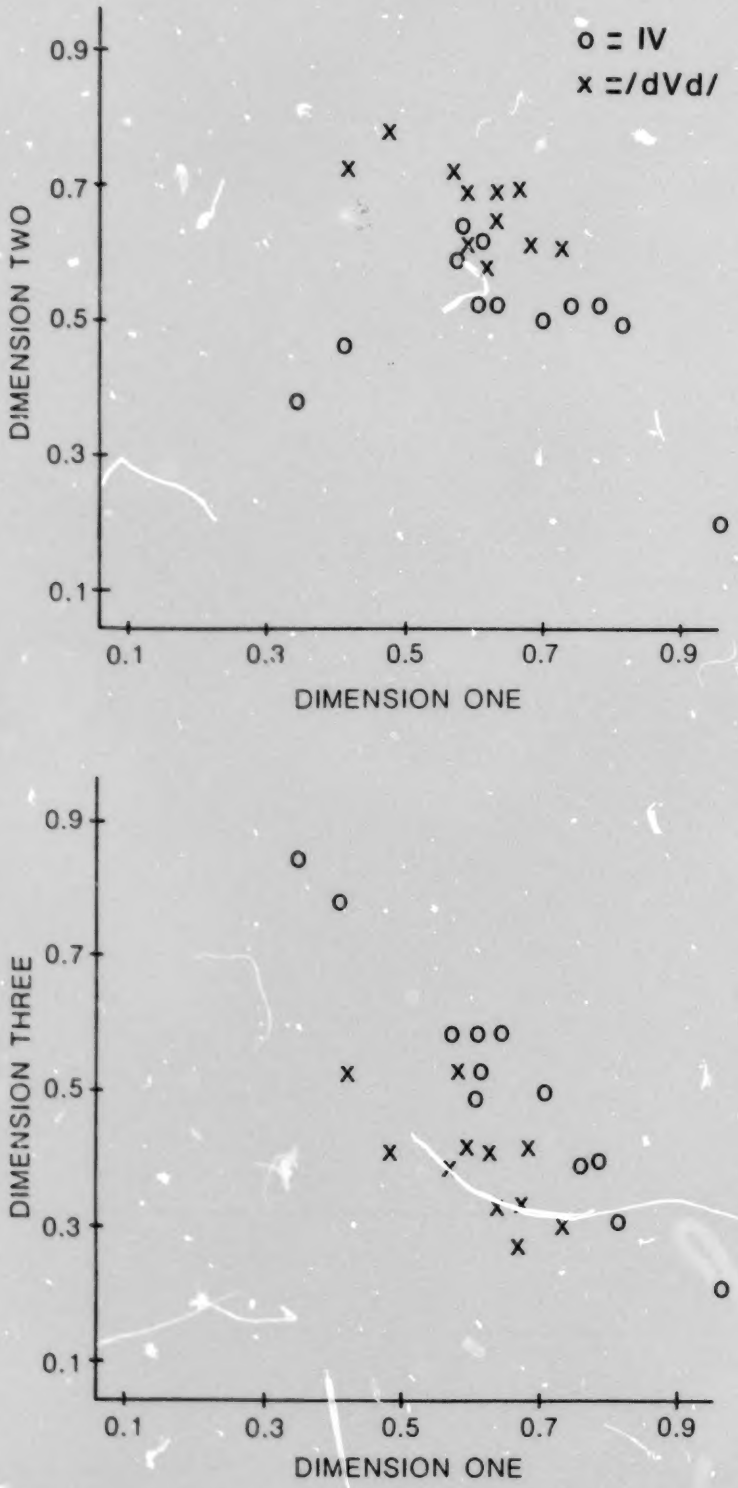


Figure 4. Adjusted weight space for Experiment 1.

Discussion

How is this effect of consonantal context to be understood? It will be argued that, broadly speaking, there are two classes of accounts that might be brought to bear to explain it and that the results of Experiment 2 lend at least suggestive support to one over the other. The first class, which will be called knowledged-based accounts, turns on a subject's understanding of certain regularities in the occurrence of vowel categories. The second class, called stimulus-based accounts, turns on a subject's sensitivity to properties of the stimuli themselves. To illustrate the differing character of knowledge-based and stimulus-based accounts, several examples of each are provided below.

Knowledge-based accounts. One plausible example of a knowledge-based account is motivated by the fact that in English vowels most generally occur in some consonantal context. The context condition of the present experiment might therefore be expected to engage linguistic processing most effectively. Such an argument is encouraged by the observation that frequency of occurrence does positively affect performance on a number of other indices to language skill, such as the reaction time to identify a word as being in one's lexicon (Forster & Chamber, 1973). On the other hand, a wealth of linguistic phenomena resist explanation in such terms. Witness, in this regard, the fact that readers of Japanese name colors more rapidly when they are written in kana (a representation of the phonologic structure of the language) than in kanji (a logographic representation) even though the latter form is seen much more often (Feldman & Turvey, 1980).

A knowledge-based account of a rather different sort could draw on the fact that certain phonological rules for vowel usage are specific to consonantal context. One such rule that has already been mentioned is that "tense" vowels can occur at the ends of syllables but "lax" vowels cannot. A listener asked to make judgments about the linguistic quality of vowel sounds might therefore have a difficult time with an isolated "lax" vowel like /I/, since no such isolated vowel is allowed in English. Singh and Woods (1970) advanced just such an argument to account for the fact that they found no evidence that tenseness had perceptual significance for listeners who rated the relative similarity of a set of isolated American English vowels. On the basis of the present findings, however, that failure might possibly be attributed to the fact that those investigators averaged their data over subjects prior to scaling. For certain isolated-vowels subjects in the present experiment, the tenseness dimension was particularly salient. For others, however, it had little or no salience. Averaging over all subjects, then, could "wash out" any statistical evidence of the significance of the tenseness feature.

Investigators (Assman et al., 1982; Macchi, 1980) have also pointed to this phonological restriction on isolated vowel usage as a potential explanation for the recurring observation that vowels can be identified more accurately in consonantal context than out (e.g., Gottfried & Strange, 1980; Strange et al., 1976; Strange et al., 1979). This cannot explain the phenomenon in full, however, since Strange et al. (1979) have also observed a consonantal influence in CV syllables (e.g., "be") in which "lax" vowels are as phonologically impermissible as they would be in isolation.

Whatever the outcome of these individual debates, the tenor of this sort of knowledge-based account is clear: to the degree that listeners are sensitive (either consciously or unconsciously) to the fact that a phonological rule of English prescribes the occurrence of certain vowel sounds in isolation, those listeners' linguistic judgments may be affected.

Recently, it has been shown that a knowledge of how speech sounds are written may have an effect as well. For instance, listeners will more rapidly detect the rhyming quality of spoken words when those words are spelled alike (e.g., "fight"/"right") than when they are not ("you"/"two") (Seidenberg & Tanenhaus, 1979). It is perhaps relevant, then, to note that the subjects in the present experiment were literate and therefore had had a great deal of experience in reading and writing vowels. In at least one previous study (Diehl et al., 1981) it has been suggested that such experience can lead to a perceptual bias in favor of consonantal context.

Knowledge-based accounts of the consonantal influence have in common the fact that they look to a subject's long-term experience with stimuli of a particular type. Such accounts would have it, for example, that extended acquaintance with frequently occurring items, or with certain phonological or orthographic regularities regarding those items, explains the perceptual effect that was observed in Experiment 1. Thus, the "locus" of knowledge-based effects is at some remove from immediate stimulation. That is to say, these accounts have much more to do with the sorts of accumulated knowledge that might be brought to bear in processing stimulus information than they do with the information itself. Not so the more stimulus-based accounts that will now be considered.

Stimulus-based accounts. As examples of stimulus-based accounts, consider two that are motivated by the fact that (as is typically the case) the isolated-vowels stimuli tended to exhibit relative spectral constancy over their course (only the vowels /o/ and /u/ were noticeably diphthongized), while the vowels in /dVd/ context tended to be marked by more or less continuous formant frequency change.

One reason why formant change may have been the source of the enhanced linguistic processing for vowels is that its presence or absence may have differentially affected the duration of a vowel's representation in what have been called auditory and phonetic memory stores. By hypothesis, the former preserves a relatively unprocessed "neural analog" or the acoustic signal and the latter preserves features of the input that are specifically relevant to speech. Fujisaki and Kawashima (1969, 1970; see also Pisoni, 1973) have pointed to the differential presence of vowels and consonants in these memories as a potential psychoacoustic basis for the observation that the former (particularly isolated vowels) tend to be less categorically perceived than the latter (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973). The present argument would simply extend this reasoning to a perceptual difference that holds between two classes of vowels, those in and out of consonantal context.

An alternative reason why formant change might be expected to engage linguistic processing particularly effectively is that properties of the speech signal are, most generally, dynamic in character (Liberman, 1982). This is a consequence of the fact that the several segments of an utterance tend to impose competing demands on the articulators, making it necessary for

talkers to interleave their productions in a series of rapid articulatory gestures. By the laws of physical acoustics, these gestures result in an assortment of dynamic modulations of the signal. Owing to this fact, a speech perceiver might be expected to be particularly attuned to any sort of acoustic change.

Though other stimulus-based accounts might be advanced, it should be apparent from just these that the focus of all such accounts will be on some acoustic property or properties of the stimulus set.

How to distinguish between the two classes of accounts. An essential difference between knowledge-based and stimulus-based accounts has to do with the degree to which they reflect a sensitivity to properties of the stimulation. Since knowledge-based factors turn on long-term experience with stimuli of a type, they should be relatively little affected by the immediate experience obtained through any particular encounter with a stimulus. Stimulus-based factors, by contrast, are expressly defined in terms of stimulus properties. It should be possible, then, to gain some evidence as to the "locus" of the consonantal influence observed here by looking at a case in which the relative contribution of immediate stimulation is reduced. If knowledge-based factors were critical to the present result, they should be expected to be manifest there as well. If, instead, stimulus-based factors were most important here, the consonantal influence should be diminished. Experiment 2 provides a case relevant to these predictions.

Experiment 2

In this experiment, subjects' memories for vowel sounds were examined with a procedure analogous to that employed in Experiment 1. The subjects were asked to imagine vowels—as occurring in isolation or in /dVd/ context—and to make judgments about the linguistic relationships among the images. It was expected, first of all, that an analysis of these judgments might help to clarify the results of Experiment 1. There it was found that the presence of consonantal context had the effect of evoking somewhat more linguistic perceptual processing of vowels than occurred in its absence and it was concluded that while a number of different accounts of this effect could be put forth, broadly speaking, these either turned on various properties of the stimuli themselves or on properties having more to do with the occurrence and recurrence of vowels as meaningful categories in English. If these latter, more knowledge-based factors are the critical ones, then it might be expected that the presence or absence of consonantal context would affect the outcome of this memory experiment no less than it did that of the perception experiment, because vowel usage rules, orthographic regularities of vowel transcription, and so on remain in force here. On the other hand, to the degree that stimulus properties are critical to the effect, the condition difference should be reduced here (or perhaps eliminated) since vowel memory is at some remove from the acoustic stimulation.

The results of this second experiment could also prove useful in a second way: they could point to the dimensions of organization for subjects' long-term memorial representations of vowels. A question arises, for instance, about the number of such dimensions. Are there three as in Experiment 1, and if so, are these the same three linguistically-meaningful dimensions that were found to have perceptual import? And it may be asked whether the same dimensions are utilized in the same way by different subjects.

Method

Stimuli. The stimuli for Experiment 2 were written analogs of the spoken stimuli used in Experiment 1. That is to say, they were orthographic representations of both isolated vowels and vowels in a trisyllabic frame in which the medial syllable was stressed /dVd/. The stimulus set comprised the same ten vowels that were used in the perception experiment. Table 2 presents a summary of all stimuli.

Table 2

Stimuli for Experiment 2. IV = isolated vowels, /dVd/ = vowels in consonantal context.

	"Spellings"		English Exemplars		
Vowel	IV	/dVd/	1	2	3
i	EE	ADEEDA	eat	heel	brief
ɪ	IH	ADIHDA	it	him	brim
ɛ	EH	ADEHDA	egg	hen	bread
æ	AE	ADAEDA	at	ham	brash
ʌ	UH	ADUHDA	up	hull	brush
ɑ	AH	ADAHDA	odd	hop	bronze
ɔ	AW	ADAWDA	ought	haul	brawn
o	OH	ADOHDA	oat	home	broach
ʊ	UU	ADUUDA	oomph	hood	brook
u	OO	ADOODA	ooze	hoop	broom

In English orthography there are numerous ambiguities with respect to the spelling of vowel sounds. The letters "OO," for example, stand for the vowel /u/ in the word "tool" and for /ʊ/ in the word "book." There are indications that these spelling ambiguities can affect listeners' perceptions of vowels (Assmann et al., 1982) and, while the present experiment was not strictly perceptual, it was thought advisable to devise vowel spellings that were unique to each sound. These are presented in the second and third columns of Table 2. In all cases the vowels were spelled with two-letter sequences. These sequences were presented alone for isolated vowels and embedded in the frame AD__DA for the trisyllables. In the latter case subjects were told to read each stimulus as a three-syllable nonsense word, the first and last syllables of which consisted of unstressed schwa (/ə/) vowels and the middle syllable of which was a stressed /dVd/.

Subjects were familiarized with the new orthography with the aid of a training sequence. This sequence paired each written vowel form with three English monosyllabic words containing the vowel sound that the form was meant to represent (see Table 2). These words were selected so as to be similar to, but distinct from, both the isolated and /dVd/ contexts employed in the experimental test.

The test series consisted of triads of stimuli presented in three adjacent columns. All possible triadic combinations of the vowels were included in each series. The order of occurrence of triads was randomized, as was the assignment of the words of each triad to the columns.

Procedure

Instructions. As nearly as possible, the instructions for Experiment 2 paralleled those for Experiment 1. It was explained to subjects that their task would be to imagine a number of different vowel sounds and to make linguistic comparisons of the images. A sense of what it would mean to make linguistic comparisons was again provided by the example of distinguishing an adult's and a child's productions of the vowel /i/ from their productions of vowels such as /ɛ/ and /ɪ/ (see the Instructions section of Experiment 1).

The triadic comparisons testing procedure was explained in detail. Subjects were told that they would be given a test series (either the isolated-vowels series or the consonantal-context series) and a cover sheet. The cover sheet had a small slit cut in it to allow the viewing of only one test line (a trial) at a time. The procedure was: (1) to move the cover sheet down the test page, thereby exposing the three stimuli of a trial; (2) to make a triadic comparison among the images of the three vowels represented on the line; (3) to write down the column numbers of the most-alike and least-alike vowel pairs; and (4) to proceed to the next trial. It was emphasized to subjects that they were under no time pressure. To the contrary, they were instructed to proceed at whatever pace they found comfortable, with the constraint that they not look back at any trial once it had been completed.

Orthographic training and administration of a pretest. Prior to the test, subjects were given extensive work with the orthographic training sequence. The experimenter first read the sequence aloud, pointing out potential errors to be avoided. Next, subjects were allowed to ask questions about the spellings and then the sequence was read aloud a second time. Finally, the subjects were told to study the sequence on their own for as long as was needed to commit the spellings to memory.

At the end of the individual study sessions and before the actual test series were presented, the subjects were asked to complete a pretest designed to assess competency with the new orthography. The pretest was straightforward: Subjects were presented a randomized list of written vowel stimuli and asked to give three examples of English words that contained the vowels indicated. The examples they gave had to be different from those used in the training sequence. Subjects' test results were omitted from the data analysis if they made more than one error on the pretest.

Subjects. Thirty-three undergraduates, enrolled in an introductory psychology course at the University of Connecticut, participated in the experiment for course credit. These individuals were native English speakers. They had no prior knowledge of either the purpose of the experiment or its design. On the basis of their performance on the pretest, six subjects were eliminated. Twelve of the remaining 27 subjects were in the isolated-vowels condition, 15 were in the consonantal-context condition.

Analysis of the Data

Since Experiment 2 was designed to test subjects' memories for vowels in a way that paralleled the perception test of Experiment 1, there was reason to expect that the most appropriate scaling solution for the present data might be the three-dimensional one that appeared earlier. After various modeling alternatives were examined, it was concluded that, with one methodological exception to be noted below, this was in fact the case.

Dimensionality of the space. The percentage of variance accounted for was computed as a function of the number of modeling dimensions. This function had roughly the same shape as its counterpart in Experiment 1, an observation consistent with the expectation that a three-dimensional modeling of these data might prove as appropriate here as it did in Experiment 1. The VAF comparison with Experiment 1 also showed that at each dimension level these memory data were somewhat less well fit by the model than were the perception data, which is to say that the data were somewhat "noisier" here. This is not surprising. In the perception test, the items presented to subjects were highly familiar (spoken English vowels) and were, in fact, the perceptual objects of interest. Here, by contrast, the items presented were rather unfamiliar vowel spellings, which only mediated contact with the memory images that were the true objects of study.

Nonmetric scaling. It has been pointed out that the stability of a modeling outcome must be considered when making decisions about scaling (see Wish & Carroll, 1974, for a discussion of this point). With respect to the present study, this consideration bore most directly on the decision to perform nonmetric individual-differences scaling, as against a more commonly used metric procedure such as INDSCAL (Carroll & Chang, 1970). For certain of the analyses of Experiment 1 in particular, the metric/nonmetric modeling distinction made little or no difference in the outcome.[8] However, this could not be said to be the case in Experiment 2; modeling these data at the metric scale resulted in an uninterpretable group space. At the nonmetric scale, on the other hand, the group space was not only interpretable, but was quite evidently related to the group space of Experiment 1.

This perception/memory difference in what might be called "measurement level" may be interesting in its own right. It suggests that the memory space for vowels is a sort of nonlinearly transformed version of the perceptual space. Interval relationships among the vowels hold in perceptual space but not in memory. On the other hand, the relatively noisy character of the memory data has already been noted, and the "measurement level" difference between the perception and memory experiments may simply reflect task variables. Whatever the true state of affairs, the approach taken in this study has been to model all data at the more conservative nonmetric level.

Starting configuration. It proved to be the case that the three dimensions of the group space could not be interpreted as originally modeled. They neither corresponded to the linguistic features of advancement, height, and tenseness as they had in Experiment 1, nor to other recognized features of articulatory or acoustic description for vowels. This was equally the case for the two- and four-dimensional group spaces.

The scaling procedure was therefore rerun in three dimensions with the group space of Experiment 1 taken as a starting configuration.[9] This was, in effect, a test of the appropriateness of that earlier group space as a model for memory data structure. It did prove to be an appropriate model, as evidenced by the fact that it fit the memory data nearly as well as had the original, uninterpretable, three-dimensional solution (mean VAF was 59% with the starting configuration, 61% without it).

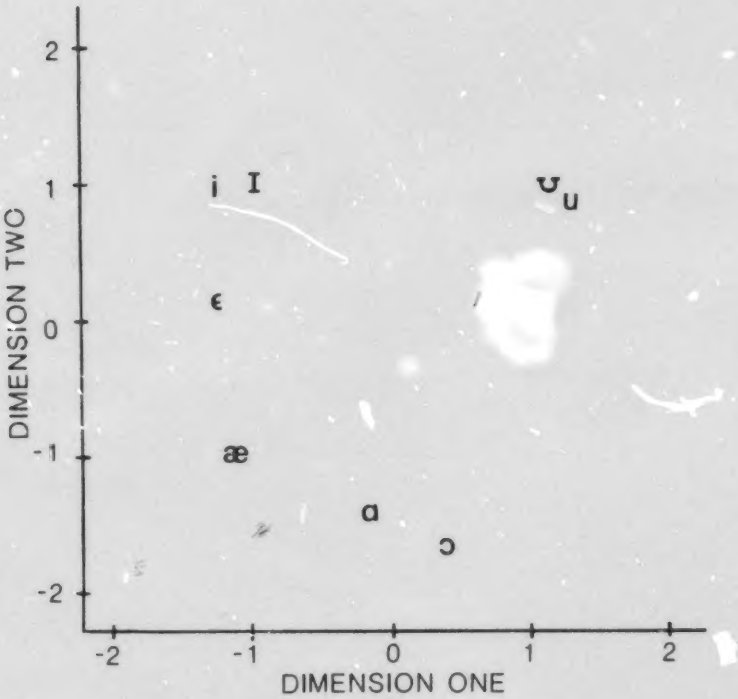
Results

Group space. The group space for all subjects who participated in the two conditions of the memory experiment is shown in Figure 5. It is quite evidently similar to the group space for Experiment 1 (Figure 2). In the dimension 2-by-dimension 1 plane, only the vowel /ʌ/ has shifted its position substantially: it is "higher" and more "fronted" in the present analysis. In the dimension 3-by-dimension 1 plane, the only vowels that moved noticeably are /æ/ and /ɑ/. The former can be seen to have taken on a dimension 3 value that is somewhat more "tense," the latter one is more "lax." These shifts do not substantially alter the overall configuration, however. On the whole, then, it can be said that this combined group space for the memory experiment does not differ substantially from that for the perception experiment. In both cases, the non-arbitrary axes of the space correspond to the linguistic features of advancement, height, and tenseness.

Weight space. Since the group spaces are similar for Experiments 1 and 2, it is interesting to see how the weight spaces compare. Indeed, a primary motivation for carrying out Experiment 2 was to determine whether the condition difference in dimension weightings seen for perception would be manifest in memory as well. A look at Figure 6, which displays the combined weight space for Experiment 2, indicates that it was not.

In Experiment 1, subjects in the consonantal-context condition were consistent with one another in attaching substantial weight to all three linguistically-meaningful dimensions of the group space, while isolated-vowels subjects were quite variable, with different individuals weighting different dimensions disproportionately. Here, by contrast, subjects in both conditions behaved in a fairly comparable way: they clustered toward the center of the weight space (roughly as did the consonantal-context subjects of Experiment 1).[10] It turned out that isolated-vowels subjects were, if anything, less variable in exhibiting this pattern than were their consonantal-context counterparts--the opposite result from that observed in Experiment 1. (This trend was not significant in the original weight space shown in Figure 6, $F(14,11) = 2.27$), but was in a weight space adjusted to compensate for goodness-of-fit differences among subjects (cf. Experiment 1, $F(14,11) = 4.30$, $p < .01$.)

Clearly, the pattern of dimension weightings obtained for memory judgments made at some remove from the acoustic stimulation is substantially different from that obtained in perception. This strongly suggests that stimulus-based factors were critical to the perceptual influence of consonants that was observed in Experiment 1.



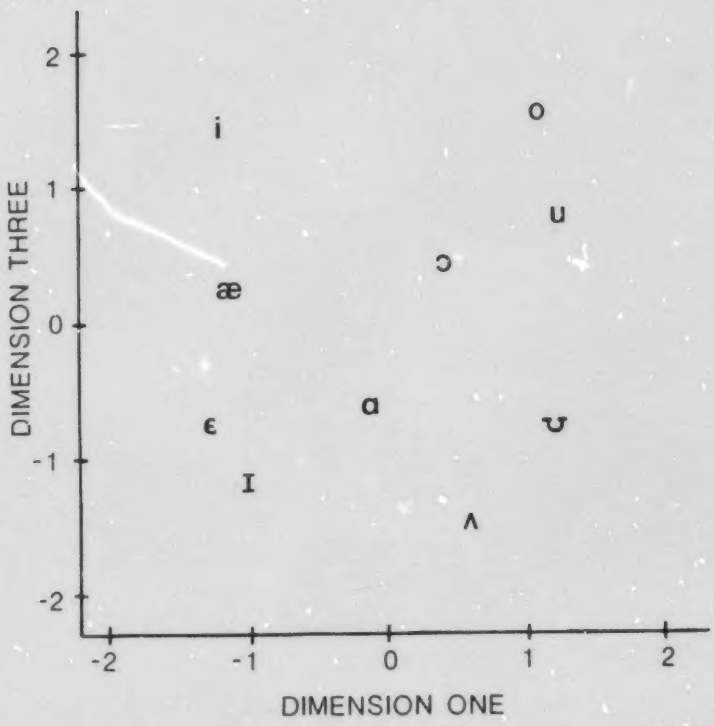


Figure 5. Group space for Experiment 2.

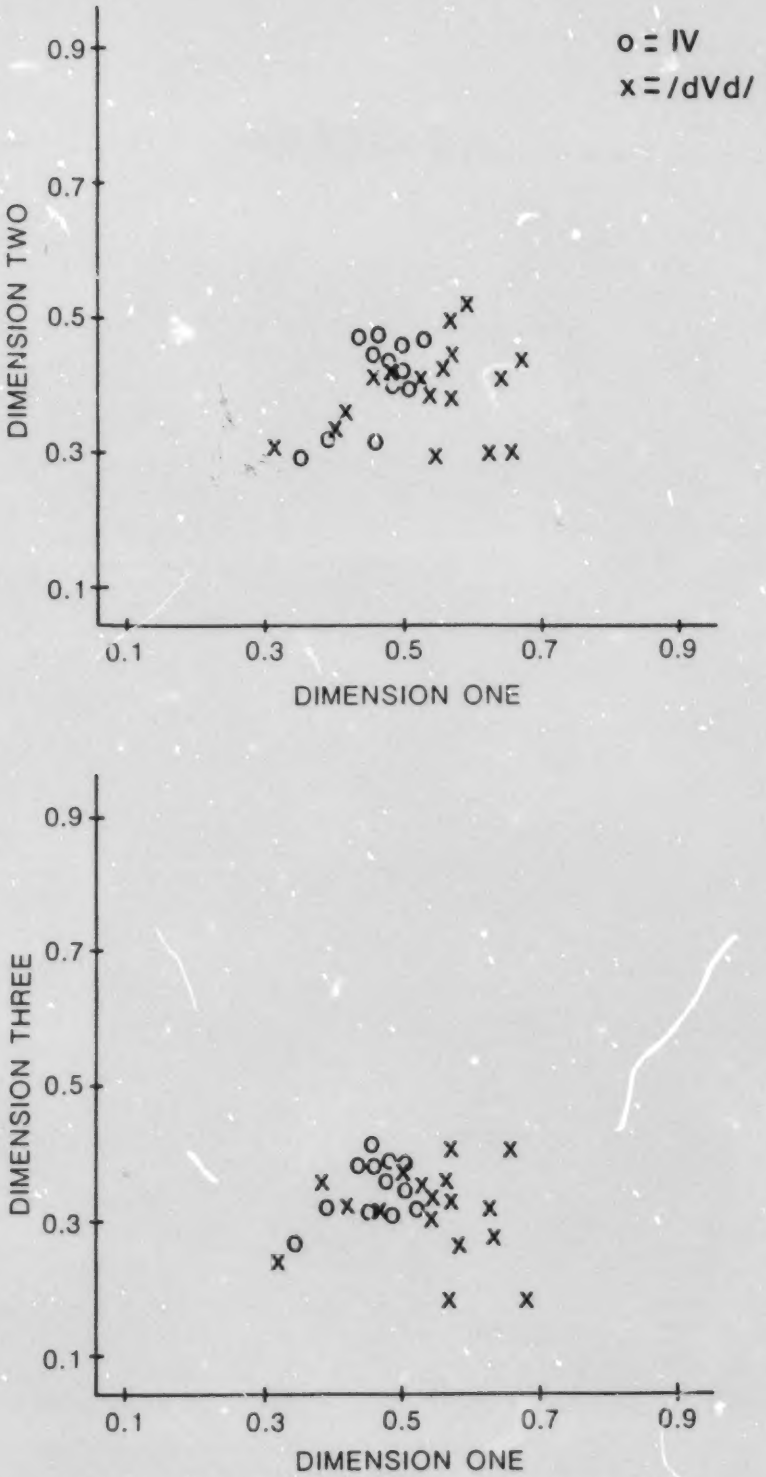


Figure 6. Weight space for Experiment 2.

Discussion

In Experiment 1, it was concluded that /dVd/ context had the effect of evoking more linguistic perceptual processing of vowels than occurred in isolation. There are a number of knowledge-based accounts of why this might have been the case, including the facts that vowels more frequently occur in consonantal context than out, that certain phonological rules are specific to consonantal context in English, and that regularities (and irregularities) in English orthographic representations of vowels may differ with context. Since these knowledge-based factors reflect a history of experience with vowels as meaningful categories in English, it might be expected that they would have an influence in this vowel memory experiment as well. However, the variance analyses of the subject weights indicate that they did not. It can be at least tentatively concluded, therefore, that the consonantal influence in perception had more to do with stimulus-based factors than with knowledge-based factors.

In Experiment 1, a close correspondence was observed between three features of linguistic description for vowels (advancement, height, and tenseness) and the three dimensions of the group space. The fact that individual-differences scaling showed these to be the dimensions that optimally accounted for variance in the several subjects' data was taken as particularly strong evidence that those linguistic features have some significance for the perception of vowels. A related point can now be made with respect to vowel memory, although it must be somewhat tempered by the reservation that the present analysis was initiated by a starting configuration. The orientation of axes for the resulting group space was nevertheless dictated by the character of individual subject data, and the observed correspondence between the linguistic features and the dimensions of this space strongly suggests that listeners' memories for vowels are, at least in some measure, organized in a way that respects those features. Thus, there appears to be a consistent recurrence of the features in perception, memory, and in linguistic behavioral data such as those having to do with grammatical rules for vowel usage.

It is important to recognize that altogether different results might have obtained. First of all, the several subjects participating in this memory experiment might have exhibited no consistent pattern of responding at all, in which case the model would have failed to account for a reasonable percentage of the variance in the data and the dimensions of the group space would have been uninterpretable. Alternatively, to the degree that subjects behaved consistently, they might have done so in a way that made little or no sense from a linguistic standpoint. Since the stimuli of this experiment were presented by eye, subjects might, for example, have made their judgments on the basis of visual features of the input, but they did not.

Summary and Conclusions

This study was motivated by an interest in the question of whether vowel perception is greatly influenced by the consonantal context in which a vowel occurs. A good deal is known about the perception (and production) of isolated vowels, and an answer to this question of consonantal influence will determine how researchers generalize from that knowledge base. To the degree that the influence on perception is minor, the isolated vowel form might reasonably be viewed as canonical (since it is unencumbered by any context effects at

all) and its acoustic signature might be taken to be composed of the essential information for vowel perception. On the other hand, if consonantal context is found to affect the perception of vowels significantly, then the isolated form can only be considered to be one variant of the vowel and, given the infrequency of its occurrence in natural speech, an arguably unrepresentative variant. Caution would therefore be required in generalizing from what is known about it.

The results of the present study clearly support the latter position. Vowels were here found to be perceived significantly differently in consonantal context than they were in isolation. One aspect of that difference was that listeners exhibited greater agreement with one another about the linguistic relationships that held among a set of vowels when those vowels were in /dVd/ context than when they were in isolation. A second aspect was that with isolated vowels listeners attended in a piecewise manner to three different vowel dimensions, while with vowels in context they integrated those dimensions in a way that was more consistent with other aspects of linguistic behavior.

These findings have been interpreted as indicating that /dVd/ context had the effect of eliciting more linguistic perceptual processing of vowels than occurred when they were presented in isolation. To the degree that this interpretation in appropriate, it follows that those who do linguistic research on vowels in the future would do well to examine them in some consonantal context.

References

Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. Journal of the Acoustical Society of America, 71, 975-989.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multi-dimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika, 35, 283-319.

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. II). New York: Academic Press.

Chiba, T., & Kajiyama, M. (1958). The vowel and its structure. Tokyo: Phonetic Society of Japan.

Diehl, R. L., McCusker, S. B., & Chapman, L. S. (1981). Perceiving vowels in isolation and in consonantal context. Journal of the Acoustical Society of America, 68, 239-248.

Feldman, L. B., & Turvey, M. T. (1980). Words written in Kana are named faster than the same words written in Kanji. Language and Speech, 23, 141-147.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior, 12, 627-635.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. Language and Speech, 5, 171-189.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute (University of Tokyo), 28, 67-73.

Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute (Faculty of Engineering, University of Tokyo), 29, 207-214.

Gottfried, T. J., & Strange, W. S. (1980). Identification of coarticulated vowels. Journal of the Acoustical Society of America, 68, 1626-1635.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 33, 469-506.

Hanson, G. (1967). Dimensions in speech sound perception: An experimental study of vowel perception. Ericsson Technics, 23, 3-175.

Hockett, C. (1958). A course in modern linguistics. New York: Macmillan.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-113.

Jakobson, R., & Waugh, L. R. (1979). The sound shape of language. Bloomington, IN: Indiana University Press.

Joos, M. (1948). Acoustic phonetics. Language, 24, Suppl., 1-136.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Beverly Hills: Sage Publications Inc.

Ladefoged, P. (1971). Preliminaries to linguistic phonetics. Chicago: University of Chicago Press.

Ladefoged, P. (1975). A course in acoustic phonetics. New York: Harcourt Brace Jovanovich.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides and diphthongs. Journal of the Acoustical Society of America, 33, 268-277.

Liberman, A. M. (1982). On finding that speech is special. American Psychologist, 37, 148-167.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A. M., & Pisoni, D. B. (1977). Evidence for a special speech-perceiving subsystem in the human. In T. H. Bullock (Ed.), Recognition of complex acoustic signals (pp. 59-76). Berlin: Dahlem Konferenzen.

Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 35, 1773-1781.

Macchi, M. J. (1980). Identification of vowels spoken in isolation and in consonantal context. Journal of the Acoustical Society of America, 68, 1636-1642.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24, 175-184.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 13, 253-260.

Pisoni, D. B. (1979). Does a listener need to recover the dynamic vocal tract gestures of a talker to recognize his vowels? Journal of the Acoustical Society of America, 65, S6. (Abstract)

Pols, L. C. W., van der Kamp, L. J. Th., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. Journal of the Acoustical Society of America, 46 458-467.

Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. Journal of the Acoustical Society of America, 53, 1093-1101.

Rakerd, B. (1982). Vowels in consonantal context are perceived more linguistically than isolated vowels: Evidence from an individual differ-

ences scaling study. Unpublished doctoral dissertation, University of Connecticut.

Rakerd, B., Verbrugge, R. R., & Shankweiler, D. P. (in press). Monitoring for vowels in isolation and in a consonantal context. Journal of the Acoustical Society of America.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. Journal of Experimental Psychology: Human Learning and Memory, 5, 546-554.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika, 27, 125-140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. Psychometrika, 27, 219-246.

Shepard, R. N. (1972). Psychological representation of speech sounds. In E. D. David & D. P. Denes (Eds.), Human communication: A unified view. New York: McGraw Hill.

Singh, S., & Woods, D. R. (1970). Perceptual structure of 12 American English vowels. Journal of the Acoustical Society of America, 49, 1861-1866.

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. Journal of Speech and Hearing Research, 6, 111-128.

Strange, W., Edman, T. R., & Jenkins, J. J. (1979). Acoustic and phonological factors in vowel identification. Journal of Experimental Psychology: Human Perception and Performance, 5, 643-656.

Strange, W., & Gottfried, T. L. (1980). Task variables in the study of vowel perception. Journal of the Acoustical Society of America, 68, 1622-1625.

Strange, W., Verbrugge, R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. Journal of the Acoustical Society of America, 60, 213-224.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 42, 7-67.

Terbeek, D., & Harshman, D. (1971). Cross-language differences in the perception of natural vowel sounds. UCLA Working Papers in Phonetics, 19, 26-38.

Wish, M., & Carroll, J. D. (1974). Applications of individual differences scaling to studies of human perception and judgment. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. II). New York: Academic Press.

Footnotes

[1]The "appropriateness" of point positioning has to do with the distances between the points. Those distances should, as nearly as possible, be ordered in a manner that reflects order in the perceptual data (see also Footnote 9).

[2]With other scaling methods, such as those designed for the analysis of single matrices of data (e.g., Shepard, 1962a, 1962b; Guttman, 1968), it is necessary to perform a post hoc rotation of the scaling solution in order to bring it into any sort of interpretable orientation. The particular rotation performed is necessarily shaped by an investigator's intuitions about the appropriate dimensions of interpretation and is, correspondingly, vulnerable to

the challenge that some other dimensions would have been equally (or more) appropriate had some other rotation been carrried out. It is just this post hoc rotation that is precluded with individual differences scaling (Carroll & Chang, 1970; Wish & Carroll, 1974).

[3]Since this is the most commonly discussed index to fit, it is the one that will be considered here. Nevertheless, it should be noted that the data were in fact scaled with a procedure (ALSCAL, designed by Takane et al., 1977) that, for computational reasons, optimizes a related but slightly different index called SSTRESS. This undoubtedly accounts for the slight decrement in the VAF function seen at five dimensions (there was no such decrement in the SSTRESS function). Solutions obtained by optimizing SSTRESS are extremely similar to those obtained with alternative individual differences scaling methods (Takane et al., 1977).

[4]VAF might not increase with an increase in dimensionality if a scaling algorithm was halted after a fixed number of iterations or, more commonly, if it was halted due to the encounter of a "local minimum" in the optimizing function (see also Footnote 3).

[5]In English, the tenseness and length feature labels are not quite so interchangeable as are, say, height and compactness. The "tense" vowels are generally "long" vowels as well, but there is one notable exception: the vowel /æ/. This vowel is phonologically "long," yet a usage rule treats it as "lax" in that it cannot appear in open position. With respect to the group space, /æ/ is likewise grouped with the "lax" vowels along dimension 3, which makes the choice of the tenseness label particularly appropriate for this dimension.

[6]The Pythagorean theorem holds that the distance between two points in a three-dimensional space will be equal to the square root of the sum of the squared distances between those points' coordinates along the three reference axes. Hence, the distance between subject 1 (indexed by the coordinates x_1, y_1, z_1) and the centroid for all subjects in the same condition (indexed by x_c, y_c, z_c) was computed with the equation:

$$\text{distance} = ((x_1-x_c)^2 + (y_1-y_c)^2 + (z_1-z_c)^2)^{1/2}$$

[7]It is also noteworthy that when the [illegible] conditions of the experiment were modeled <u>separately</u>, this significant difference in subject variability remained (see Rakerd, 1982, for the analysis).

[8]In Experiment 1, the group and weight spaces for the metric analysis of the combined data were virtually identical to those for the nonmetric analysis shown in Figures 2 and 3. When the conditions were modeled separately (Rakerd, 1982), the metric and nonmetric solutions did differ, at least in detail.

[9]To get a sense of what employing a starting configuration entails, it is important to understand how the scaling procedure operates more generally. An optimal fit to a set of data is achieved by successively adjusting the stimulus configuration over a series of iterations. The scaling procedure halts when the improvement achieved on any given iteration is less than some specified tolerance value. The adjustment that is made to the configuration amounts to moving the individual stimuli around in the group space in a way

that is sensitive to the modeling shortcomings of the existing configuration. If, for example, the vowels, /i,ɪ,ɛ/ were currently positioned in the space such that the distances among them were ordered as follows:

/i/-----/ɪ/----------/ɛ/

and yet most subjects ranked vowel-pair similarity such that /i-ɪ/ was judged less similar than /ɪ-ɛ/, then on the following iteration of the procedure there would be a shift in the positioning of /ɪ/ to correct the mismatch between model and data, i.e.:

/i/----------/ɪ/-----/ɛ/

Since the scaling procedure is capable of making such adjustments, it is possible to start with a truly random stimulus configuration and gradually, over trials, to move to one that fits a data set quite well. It is equally possible, however, to start with a configuration that, for a priori reasons, might be expected to fit the data closely from the outset. To the degree that it does, then the procedure will make only minor improvements and will halt in a relatively small number of iterations (because those improvements will be less than the halting tolerance level). Owing to this feature, it is possible, in effect, to test out the appropriateness of a particular starting configuration for the individual differences modeling of any set of data.

[10]Since scaling accounted for a relatively smaller percentage of the variance in the memory data than it accounted for in the perception data (Experiment 1), all weights are, on average, smaller here (i.e., closer to the origin of the weight space).

CHILDREN'S PERCEPTION OF [s] AND [ʃ]: THE RELATION BETWEEN ARTICULATION AND PERCEPTUAL ADJUSTMENT FOR COARTICULATORY EFFECTS

Virginia A. Mann,+ Harriet M. Sharlin++, and Michael Dorman+++

Abstract. When synthetic fricative noises from an [ʃ]-[s] continuum are followed by [a] and [u], adult listeners perceive fewer instances of [ʃ] in the context of [u] (Mann & Repp, 1980). This perceptual context effect presumably reflects adjustment for the coarticulatory effects of rounded vowels on preceding fricatives, and thus implies possession of tacit knowledge of this coarticulation and its consequences. To determine the role of articulatory experience in the ontogeny of such knowledge and the consequent perceptual adjustment, the present study examined the effect of [eɪ] and [u] on the perception of [s] and [ʃ] by children who can and cannot produce these consonants. The stimuli comprised synthetic frication noises from an [ʃ] to [s] continuum adjoined to periodic portions excerpted from natural tokens of "shave" and "shoe." The subjects included adults, five- and seven-year-old children who correctly produce both [ʃ] and [s], and seven-year-old children who misarticulate both fricatives. All three groups of children showed a significant context effect equivalent to that of adults and independent of age and the fricative articulation. Therefore, productive mastery of [s] and [ʃ] is not responsible for children's perceptual adjustment to vowel rounding on the spectra of voiceless fricatives.

Introduction

Among adult subjects, context effects in the perception of spoken consonants are a well-established phenomenon (see Repp, 1982, for a recent review). One acoustic pattern may support different phonetic interpretations in different environments. Examples of such effects can be found in the perception of bursts as cues for stop consonant place of articulation (Liberman, Delattre, & Cooper, 1952), and in the perception of formant transitions as cues to consonant place (Mann, 1980; Mann & Repp, 1981) and manner (Miller & Liberman, 1979). Another example, and the one that concerns us here, involves the place of articulation of voiceless fricative noises: When a synthetic fricative

---

+Also Bryn Mawr College.
++Bryn Mawr College.
+++Arizona State University.
Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR05596 to Haskins Laboratories. Some of the results were reported at the 102nd Meeting of the Acoustical Society of America in Chicago, May 1982. We thank Deborah Strawhun for conducting pilot stages of Experiment 1, and Jocelyn Jones for her assistance with Experiment 1.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

noise ambiguous between [ʃ] and [s] precedes the vowel [u], listeners perceive [ʃ] less often than when the same noise precedes the vowel [a] (Fujisaki & Kunisaki, 1978; Mann & Repp, 1980).

Like a myriad of other context effects in speech perception, the contrasting effect of [u] and [a] on perception of a preceding fricative noise finds a parallel, and a plausible explanation, in the dynamics of articulatory gestures and their acoustic consequences. The parallel is that, due to coarticulation of adjacent phonemes, when [ʃ] and [s] precede a rounded vowel, such as the English [u], they are influenced by anticipatory liprounding. The effect is a lowering of fricative noise spectra relative to that which occurs when [ʃ] and [s] are produced before an unrounded vowel, such as the English [a] (Bondarko, 1969; Heinz & Stevens, 1961; Mann & Repp, 1980). The explanation is that, since [s] noises, in general, involve higher spectral frequencies than [ʃ] noises, any compensation for the consequences of liprounding during fricative production would make a given noise appear relatively higher when it occurs before a rounded vowel, thus decreasing the likelihood that [ʃ] will be perceived.

Therefore, the tendency of adult listeners to give fewer [ʃ] responses when synthetic fricative noises occur in the context of [u] is interpreted as the reflection of a tendency to compensate for the acoustic consequences of anticipatory liprounding on fricative noise spectra (Mann & Repp, 1980). That they so take account of the acoustic consequences of articulatory dynamics as they assign phonetic labels to speech stimuli is not a unique attribute of fricative perception, but would seem to be a more general and fundamental property of perception in the speech mode. It is as if speech perception is guided by some tacit knowledge of the diverse acoustic consequences of articulatory gestures (Repp, Liberman, Eccardt, & Pesetsky, 1978), and of the subtle changes that necessarily ensue when sequences of such gestures weave and overlap in fluent speech (Mann, 1980; Mann & Repp, 1981). The basis of such knowledge, however, remains unclear, as does its role in young children's speech perception. To gain insight into these issues, the present study has explored the effects of [eɪ] and [u] on the perception of the [ʃ]-[s] distinction among children who can produce [s] and [ʃ], and those who cannot.

It is possible that tacit knowledge about the articulation of a given phoneme, and its diverse acoustic consequences, is gathered from listening to one's own production of that phoneme. If so, experience with the articulation of [s] and [ʃ] might be critical to any articulatory knowledge that allows the child to compensate for the effects of liprounding on fricative noise spectra. This hypothesis would be verified were we to find the normal contrasting effects of [u] and [eɪ] only in the perception of fricatives by children who can produce [s] and [ʃ], and not in that of children who have yet to produce these phonemes.

On the other hand, it is likewise possible that children who cannot produce [s] and [ʃ] could nonetheless be just as capable (or incapable, as the case may be) of perceptually adjusting for the influence of liprounding on fricative noise spectra. On finding this to be the case, we could reject a hypothesis that correct fricative articulation is essential to knowledge about the consequences of fricative-vowel coarticulation, and then turn to considering three alternative bases of that knowledge. First, any tacit knowledge underlying the effect of vocalic context on fricative perception might be instantiated by more general experience with one's own articulation as opposed

to specific experience with fricative articulation. Second, it could be brought about by experience with hearing and seeing the speech of others. Third, given the many firdings that at least some knowledge about the acoustic consequences of articulation could be inborn (Kuhl & Meltzoff, 1982; Miller & Eimas, in press), the ontogeny of tacit articulatory knowledge could be largely under genetic control, and relatively independent of specific experience, barring the necessary role of stimulation in the emergence of genetic behaviors.

A review of the literature reveals that, while there are many studies of the ontogeny of speech perception and production, much remains to be learned about fricative perception, and its relation to fricative production. Prelingual infants have been reported to be capable of discriminating synthetic tokens of [seɪ] and [ʃeɪ] (Eilers, 1980; Eilers & Minifee, 1975) and six-month-old infants may distinguish natural tokens of [s] and [ʃ] in the context of [a] and [u] (Kuhl, 1980). Yet when [s] and [ʃ] initiate natural CVC syllables, children aged ten to eighteen months may fail to make a perceptual distinction (Garnica, 1971; Shvachkin, 1973) and children as old as five years of age may show confusions among natural tokens of [s] and other fricative consonants (Abbs & Minifee, 1969). Likewise, although there are reports that children as young as two or three years old may correctly produce [s] and [ʃ] (Prather, Hedrick, & Kern, 1975), there is much evidence that fricatives are produced relatively late in language development, and that fricative misarticulation can be present well into the early elementary grades (Moskowitz, 1975) with considerable individual variability (Ingram, Christensen, Veach, & Webster, 1980). In short, it is unclear exactly when the [s]-[ʃ] distinction is mastered either in perception or production, nor is the relation between the two abilities apparent. On the basis of the common observation that development of language comprehension precedes that of language production, it might be tempting to discard a hypothesis that mature production of the [ʃ]-[s] distinction is essential to mature perception of that distinction. Nonetheless, there are no reports that falsify this hypothesis, nor has a subtle and sensitive assessment of children's perception of fricatives been undertaken, such as might be supplied through a study using context effects.

With these considerations in mind, we conducted two experiments, each concerned with the contrasting influence of [a] and [u] on young children's perception of the [ʃ]-[s] distinction. Our methodology is drawn from that of Mann and Repp (1980), employing a continuum of synthetic fricative noises (ranging from one appropriate to [ʃ] to one appropriate to [s]) that were followed by vocalic portions from natural syllables containing the vowel [eɪ] or [u]. Their adult subjects were required to label the initial fricative of each syllable as [ʃ] or [s], and the context effect was measured in terms of the number of [ʃ] responses given in the context of each vowel. In Experiment 1, we adapt Mann and Repp's materials and their phoneme labeling task to a forced-choice picture identification task suitable for use with preliterate children, and we provide a test of these adaptations among a population of five- and seven-year-old children who have mastered production of [s] and [ʃ]. Thus we demonstrate the utility of our procedure and discern whether any marked changes in vocalic context effects occur following the mastery of fricative production. In Experiment 2, we turn to a second population of seven-year-old children who are in speech therapy because they have not mastered production of [s] and [ʃ]. In this case, our goal is to discern whether vocalic context effects are present before fricative articulation is fully mastered.

Experiment 1

Method

Subjects. All subjects were native speakers of English who had no prior experience with synthetic speech. Adults were recruited from the Bryn Mawr area and children were recruited from a local day-care center: none of them had any known organic, behavioral, emotional, or intellectual problems. In order to be considered as a potential subject, each adult had to report no known hearing or speech pathologies. Each child had to have normal hearing acuity as determined by preschool screening and to be able to produce correctly the [s] and [ʃ] in "sue," "shoe," "save," and "shave." Chosen according to these criteria, there were ten subjects at each of three age levels in Experiment 1: five-year-olds (mean age 5.6 years), seven-year-olds (mean age 7.5 years), and adults (mean age 22.4 years).

Materials. The stimuli were hybrid syllables consisting of synthetic fricative noises followed by natural vocalic portions to form two [ʃ]-[s] continua: "shoe"-"Sue" and "shave"-"save." To construct them, we began with recordings of the words "shoe" and "shave" that had been read aloud by a native male speaker of American English as part of a list of words containing initial voiceless fricatives. All utterances were digitized at 10 kHz using the Haskins Laboratories Pulse Code Modulation (PCM) system, and the single best tokens of "shoe" and "shave" were chosen for further use. The fricative noise was then removed from each of these (the fricative noise being defined as the signal portion preceding the onset of periodicity), and replaced, in turn, with each of nine digitized synthetic fricative noises created on the Haskins Laboratories OVE IIIc speech synthesizer. The synthetic noises were characterized by two steady-state poles whose center frequencies, as can be seen in Table 1, increased in eight approximately equal steps from Stimulus 1, which approximated a natural [ʃ], to Stimulus 9, which approximated a natural [s]. Noise duration was held constant at 250 ms, with a 150 ms initial amplitude rise, and a 30 ms final amplitude fall.

Table 1

Pole Frequencies of Fricative Noises (Hz)

Stimulus	Pole 1	Pole 2
1	1957	3803
2	2197	3915
3	2466	4148
4	2690	4269
5	2933	4394
6	3199	4655
7	3389	4792
8	3591	4932
9	3917	5077

For the purpose of testing perception of the test stimuli, two different magnetic tapes were prepared, a separate one for each stimulus continuum. Each tape consisted of a practice set comprising five tokens of each of the two endpoint stimuli arranged in a random order, followed by a test set comprising a randomized sequence that included five repetitions of each of the nine test stimuli along the continuum. Interstimulus interval was held constant at 5 sec.

Procedure

All testing was conducted individually at the residence (for adults) or daycare center (for children) where the subject was solicited. Each subject listened to stimuli over circumaural earphones at a presentation level of approximately 70 dB SPL. Both tapes were completed within a single session, with the order of presentation counterbalanced across subjects. For each tape, the ten items in the practice set were presented first, followed by presentation of the 45 test items. The procedure involved the subject's listening to each stimulus and then reporting his or her phonetic perception. Whereas adults gave written responses of "s" or "sh," as in the procedure of Mann and Repp (1980), children gave a two-alternative forced-choice pointing responses to pictures that corresponded to the words on the tape—"a shoe" vs. "a girl named Sue" for the [u] context, "a man having a shave" vs. "a piggy-bank in which to save" for the [eɪ] context—and their responses were transcribed by the examiner, who did not know the identity of the stimulus being presented. To accustom children to this task the experimenter showed two pictures, "tree" and "blue," before the test tape was presented and asked the child to point to the appropriate picture as she said each word aloud. When the child correctly identified five presentations of each of these two words arranged in random order, the task was repeated using pictures for "shoe" and "blue." Finally, the child was shown the pictures for the appropriate experimental task and given practice with the experimenter saying each test word aloud. When the child had touched each picture correctly on five occasions, arranged in random order, presentation of the prerecorded practice and test stimuli followed.

Results and Discussion

The data for Experiment 1 consist of labeling responses of "s" and "sh" for stimuli along each of our two experimental continua gathered directly from adults, and inferred from children's picture verification responses. We will briefly consider the data obtained with adult subjects, then proceed to a report of the results obtained with children at each age, and a brief discussion of their import.

Adults. A summary of the results obtained with the ten adult subjects appears in Figure 1, where the average percent of "sh" responses is plotted as a function of stimulus position along the fricative noise continuum, separately for each vocalic context. Solid lines represent the results obtained when fricative noises initiated a syllable containing the rounded vowel [u], and dashed lines represent those obtained when the same noises initiated the syllable containing the unrounded vowel [eɪ]. For both continua, listeners were quite consistent in their labeling of the endpoint stimuli. And, as expected, the category boundary for the labeling function obtained in the context of the unrounded vowel from "shave" occurs between stimuli 5 and 6 (at 5.2), whereas that for the unrounded vowel from "shoe" occurs between stimuli 4 and 5 (at

4.2). Thus fewer "sh" responses were given in the context of the rounded vowel, $t(18) = 3.1$, $p < .01$.

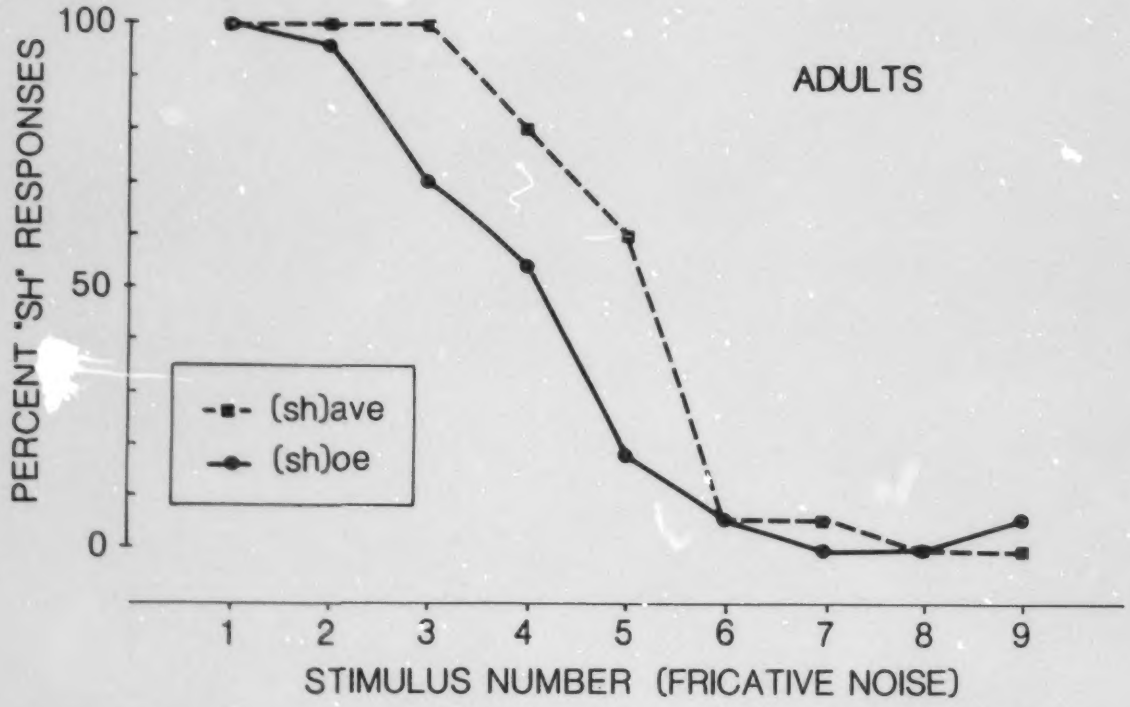


Figure 1. Influence of vocalic context on the labeling of fricative noises by adult subjects.

Children. All children successfully learned the procedure, and were 100% correct in identifying the pictures corresponding to spoken versions of the test words and 80% or better correct in responding to the practice endpoint stimuli. The results for five- and seven-year-olds are graphed in Figures 2 and 3, respectively. Here, as in the case of the adult subjects, both the endpoint stimuli were labeled quite consistently, and here, as well, the category boundaries for the two vocalic contexts lie at different locations. The boundary for noises presented in the context of the unrounded vowel lies at 5.5 for five-year olds, and 5.2 for seven-year-olds, while the boundaries for noises heard in the context of the rounded vowel occur at 4.1 and 4.3, respectively.

An analysis of variance, conducted on the total number of "sh" responses given in each vocalic context by the adults and the children at each of the two age levels, reveals a main effect of vocalic context, $F(1,27) = 59.4$, $p < .001$, but no main effect of age, and no interaction between the effects of age and vocalic context. Thus, all subjects, adults and children alike, tended to give fewer "sh" responses in the context of the rounded vowel: for five-year-olds, $t(18)=2.31$, $p < .05$, and for seven-year-olds, $t(18) = 3.37$, $p < .01$. Moreover, when measured as the difference between the number of "sh" responses given in each context, the extent of the context effect among children was not significantly different from that among adults ($p > .1$).

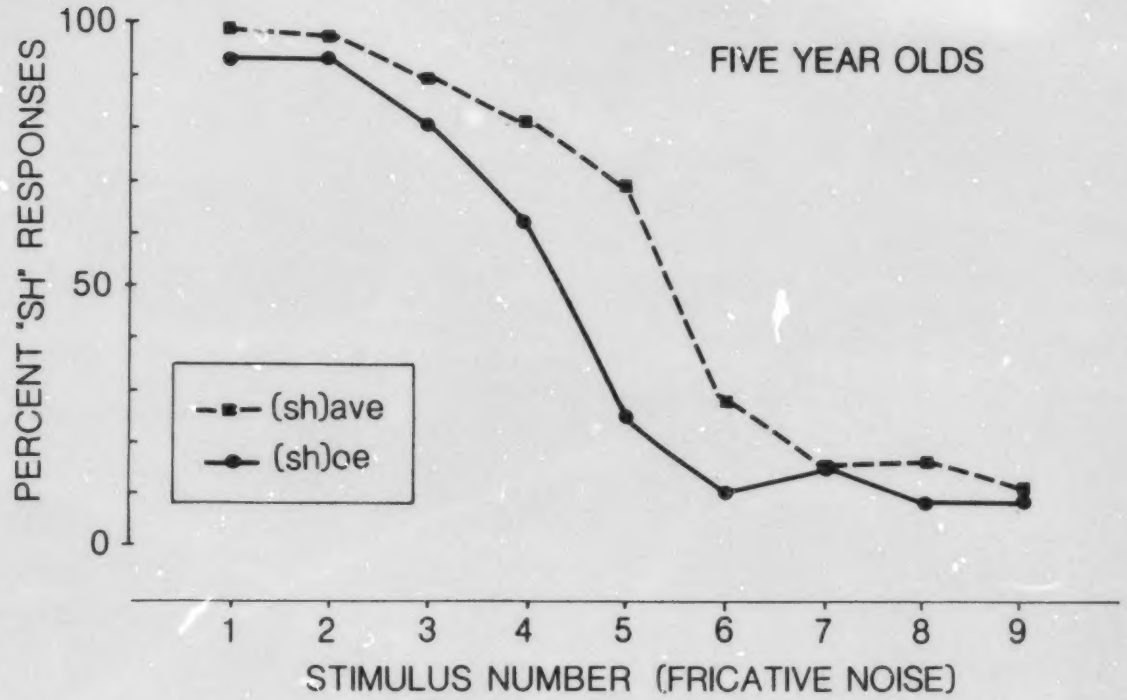


Figure 2. Influence of vocalic context on the labeling of fricative noises by five-year-olds who can articulate [s] and [ʃ].

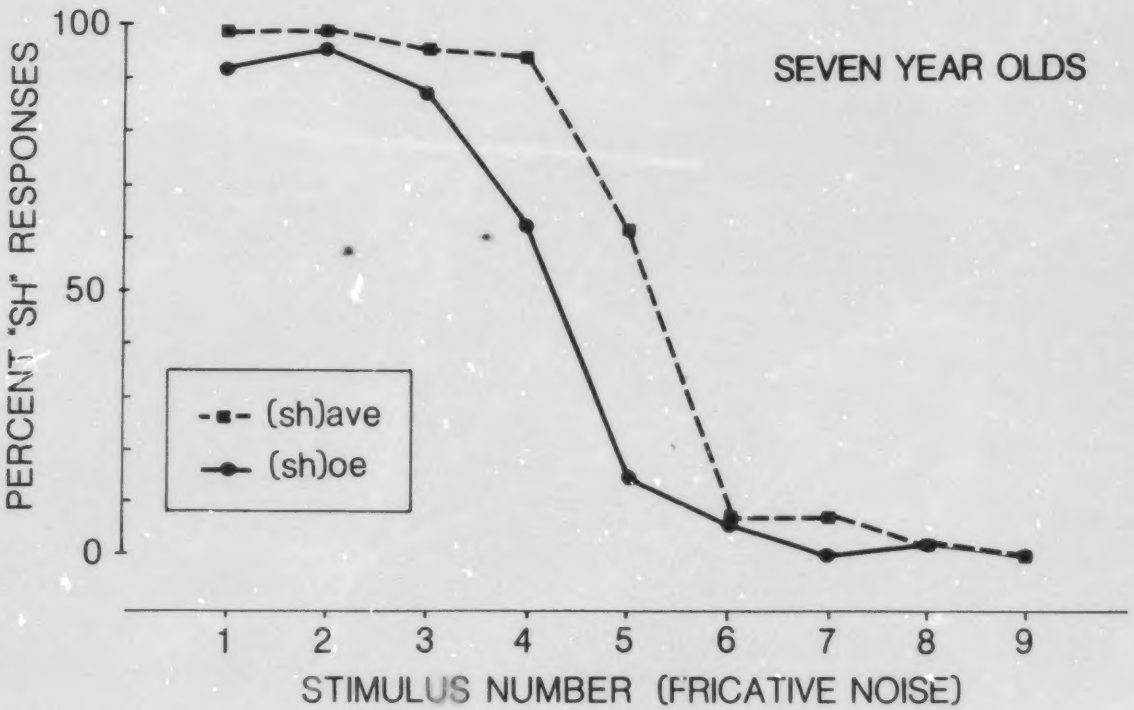


Figure 3. Influence of vocalic context on the labeling of fricative noises by seven-year-olds who can articulate [s] and [ʃ].

Using a new set of stimuli, then, Experiment 1 has confirmed previous reports (Fujisaki & Kunisaki, 1978; Mann & Repp, 1980) that when synthetic fricative noises along an [ʃ]-[s] continuum are followed by a vocalic portion that contains the vowel [u], the category boundary is shifted towards a lower noise frequency and fewer "sh" responses, than when the same fricative noises are heard in the context of the vowel [ɛɪ]. Most importantly, it has demonstrated that this vocalic context effect can be present among five- and seven-year-old children who correctly produce [s] and [ʃ], and that, among such children, the extent and direction of the effect is remarkably similar to that obtained among adults. Thus children as young as five years of age who can produce both [s] and [ʃ] show an adult-like perceptual compensation for the coarticulatory effects of liprounding on the spectra of these fricatives, and we may conclude, therefore, that knowledge of fricative-vowel coarticulation and its acoustic consequences does not markedly lag behind productive mastery of [s] and [ʃ]. Otherwise, we should have found an age-related difference between the children and adults who participated in our study. This leaves us with two possibilities as to the relation between perception and production: Either perceptual mastery precedes production mastery, or the two begin at more or less the same time. To decide between these alternatives, we turn to the second experiment of our study, which asks whether a vocalic context effect is present among children who cannot produce [s] and [ʃ].

Experiment 2

Method

Subjects. The subjects were fourteen children recruited from the second-grade classes of parochial schools in Northeast Philadelphia, who served with the permission of their parents and at the convenience of their teachers. Each of them was selected with the help of speech therapists who worked in their schools. They fulfilled all of the following criteria:

1) Incorrect production of initial [s] and/or [ʃ]; either substituting one for the other, substituting another phoneme instead, or simply omitting [s] and [ʃ] altogether.

2) No difficulty with the production of phonemes other than fricatives or affricates.

3) A maximum of one year in speech therapy.

4) Audiometry scores within the range defined in Experiment 1.

5) No soft neurological signs, cerebral palsy, emotional or behavioral disorders.

Chosen according to these criteria, there were six females and eight males, with an average age of 7.6 years.

Materials and Procedure

The materials and procedure were as in Experiment 1. Each subject was excused from his or her classroom and taken to the speech room in the school, where the experimenter explained that the child was helping her to study the way children hear language. The subjects were assured that there was no right or wrong answer involved, and that all that was required was to listen carefully. The same procedure as in Experiment 1 was used, with training followed by practice, culminating in presentation of the test trials. Order of the test tapes was counterbalanced across subjects.

Results and Discussion

The data obtained from the seven-year-old children who could not produce [s] and [ʃ] are summarized in Figure 4, which should be compared with Figures 1-3 from Experiment 1. We have combined the results across children who omitted [s] and [ʃ] (N = 8), those who substituted one for the other (N = 4), and those who substituted another phoneme instead (N = 2), as the nature of production errors did not appear to influence the pattern of results. As in the first experiment, all subjects labeled the words spoken during training with an accuracy of 100% correct, and also labeled the endpoint test stimuli with an 80% or better accuracy. Thus, they could clearly distinguish good exemplars of [s] and [ʃ]. Inspection of Figure 4 further reveals that these children also showed vocalic context effects on fricative perception. When the stimuli along the synthetic continuum were followed by the vocalic portion from "shave," the average phonetic boundary lies between stimuli 5 and 6 (5.2), whereas that for the same fricative noises followed by the vocalic portion from "shoe" lies between stimuli 4 and 5 (4.3). Thus, fewer "sh" responses were given in the context of the rounded vowel than in that of the unrounded one, $t(26) = 3.79$, $p < .005$.

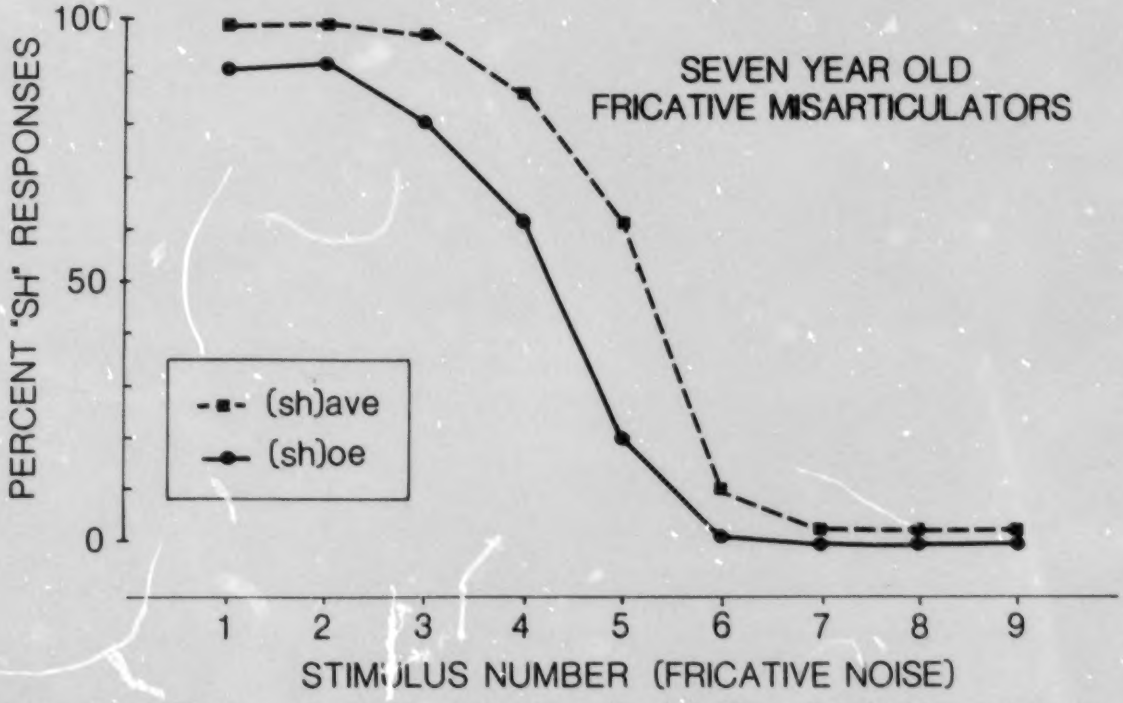


Figure 4. Influence of vocalic context on the labeling of fricative noises by seven-year-olds who cannot articulate [s] and [ʃ].

There are two points to be made in discussing these findings. The first, and most central to our concern, is that perception of [s] and [ʃ] by children who cannot produce both of these phonemes does not differ signficantly from that of adults and children of the same age who can produce them. That is, their perception of [s] and [ʃ] is affected by vocalic context in the same manner (i.e., more "sh" responses in the context of the unrounded vowel) and to an equivalent extent. Thus, it would appear that these children can take account of the consequences of coarticulation of a fricative with a following vowel, even though they do not directly control those consequences in their own speech production.

A second point, more pertinent to clinical concerns, is that the exclusive problems with fricative articulation that distinguish the children of our second experiment from those of the first experiment do not appear to be due to aberrant perceptual abilities. This is a conclusion that has been reached in several previous studies of children selectively impaired in producing liquids (Strange & Broen, 1981). Perhaps some developmental delay in motor control is the cause of selective misarticulation of fricatives and affricates, given the distinguishing developmental characteristics of this class as outlined by Ingram et al. (1980). Fricatives are avoided by many very young children, and it is not impossible that certain children merely avoid them longer than others. Also, since fricatives are among the last phonemes to be produced correctly (and there seems little agreement on the span of time involved in acquisition of other phonemes, much less this controversial class), there is ample reason to suspect that many of the children who participated in the present experiment are following a normal pattern, albeit more slowly, of phoneme acquisition.

However, before leaving this second point, we would like to recognize the possibility that certain severe articulatory problems could be based in a perceptual disorder (Strange & Broen, 1981). In this regard, we note that we have examined a group of seven-year-old children who present with multiple articulatory problems spanning three or more manner classes, and we have found them to be quite different from children who selectively misarticulate fricatives and affricates (Mann, Dorman, Strawhun, & Sharlin, 1982; Sharlin, 1982). Subjects who are multiple misarticulators give responses that tend to be more erratic; their attentiveness is also noticeably lower and they "fidget" more than the other children whom we have tested. They behave as if our task is in some way unexpectedly aversive, owing, perhaps, to an inability to competently and confidently make the required perceptual distinction. In addition, and most notably, these children are unique in their tendency as a population to show no significant effect of vocalic context on fricative perception.

General Discussion

The following general conclusions can be drawn from the results of Experiments 1 and 2: 1) Children as young as five years of age who correctly articulate [s] and [ʃ] show vocalic context effects on fricative perception that are commensurate with the context effects observed among adult subjects; 2) Competent production of [s] and [ʃ] is not necessary for the manifestation of vocalic context effects on fricative perception; 3) The exclusive misarticulation of fricative consonants, like other selective problems in speech production (see Strange & Broen, 1981), is not simply attributable to deficits in fricative perception.

We may now turn to a consideration of our findings as they pertain to the various hypotheses, outlined in the Introduction, about the source of the tacit knowledge of articulation that we hold responsible for the influence of vocalic context on fricative perception, and that we presume to be guiding mature speech perception. Certainly we may reject the hypothesis that experience with the production of fricatives is essential to the acquisition of such knowledge that allows listeners to compensate for the consequences of fricative-vowel coarticulation on fricative noise spectra. Otherwise, we should not have found vocalic context effects to be equally present in the perception of children who can and cannot produce [s] and [ʃ]. Even children who selectively omit fricatives altogether (of which we tested 8) showed vocalic context effects on fricative perception equivalent to those among other children and adults.

This leaves us to consider the remaining three possibilities about the basis of tacit articulatory knowledge. One possibility is that, while there is no simple one-to-one dependence of knowledge about the consequences of fricative-vowel coarticulation on competent production of [s] and [ʃ], some experience with language production may be essential to the acquisiton of that knowledge; for example, experience with producing rounded and unrounded vowels and observing their different consequences on sound spectra, in general. A second possibility is that tacit articulatory knowledge does not emerge through feedback from one's own articulation so much as through experience with listening to, and perhaps watching, the articulations of others. A third is that tacit knowledge is not induced by experience with one's own articulation or that of others, but is genetically given so as to be present and functioning by the age of five years, before successful fricative production (contingent, perhaps, on some type of auditory stimulation). Each of these possibilities is equally consistent with the present findings, but, as we shall now argue, they are not equally consistent with certain other findings in the literature.

Considering each possibility in turn, we note that the first is inconsistent with reports that subjects who lack speech production abilities may nonetheless demonstrate apparently normal speech perception (Forcin, 1974), and with a report by Whalen (1981) who shows vocalic context effects for non-native vowels. However, before concluding that feedback from one's own articulation is not a prerequisite for acquiring tacit knowledge about articulation and its consequences, it would be desirable to repeat the present study, using subjects with total congenital inability to speak.

We turn next to the second hypothesis, which stresses experience with the articulation of others. While this is consistent with the perceptual capabilities of inarticulate subjects, and with the late onset of certain speech perception abilities, it is at odds with findings that neonates display adult-like discrimination of many speech sounds (see, for example, Eilers, 1980, and Miller & Eimas, in press, for reviews). One might test this hypothesis by studying children who have recently been corrected for a congenital hearing loss, or by examining congenitally blind children who have not had the opportunity to observe the lip-rounding gestures of others. If subjects in these groups show normal vocalic context effects on fricative perception, it would suggest that experience with the articulations of others does not have a critical role in instantiating knowledge of articulation and its consequences. However, finding that such children fail to show context effects might be interpreted either as evidence that experience instantiates articulatory

knowledge, or as evidence that experience merely facilitates or maintains such knowledge (Gottlieb, 1976). Testing neonates and young children could ultimately decide between these two possibilities.

The third and final possibility is that tacit knowledge of articulation and its consequences is genetically endowed, as opposed to deriving from experience with the consequences of one's own articulation or with those of others. This hypothesis is consistent with findings that neonates prefer to look at a face articulating the same vowel that they are hearing (Kuhl & Meltzoff, 1982), and also display a wide range of speech perception behaviors that are directly analogous to the perceptual capabilities of adult speech perceivers, including certain context effects (Miller & Eimas, in press). It is further consistent with evidence that, although infrahuman species discriminate certain speech sounds much as human listeners do (cf. Kuhl & Miller, 1978; Waters & Wilson, 1976), and may even categorize fricative noises along an [s]-[ʃ] continuum (Sevchik, 1979), they fail to show the present vocalic context effects on fricative perception (Sevchik, 1979). If context effects that involve tacit knowledge of articulatory dynamics are unique to human listeners, then it is likely that the knowledge they depend on is genetically based.

References

Abbs, M. S., & Minifie, F. D. (1969). Effects of acoustic cues in fricatives on perceptual confusions in preschool children. Journal of the Acoustical Society of America, 46, 1535-1542.

Bondarko, L. V. (1969). The syllable structure of speech and distinctive features of phonemes. Phonetica, 20, 1-40.

Eilers, R. E. (1977). Context-sensitive perception of naturally produced stop and fricative consonants by infants. Journal of the Acoustical Society of America, 61, 1321-1336.

Eilers, R. E. (1980). Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. II, pp. 23-39). New York: Academic Press.

Eilers, R. E., & Minifee, F. D. (1975). Fricative discrimination in early infancy. Journal of Speech and Hearing Research, 18, 158-167.

Forcin, A. J. (1974). Speech perception in the absence of speech productive ability. In N. O'Connor (Ed.), Language, cognitive defects, and retardation. London: Butterworth.

Fujisaki, H., & Kunisaki, O. (1978). Analysis, recognition, and perception of voiceless fricative consonants in Japanese. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26, 21-27.

Garnica, O. K. (1971). The development of the perception of phonemic differences in initial consonants by English speaking children: A pilot study. Papers and Reports On Child Language Development, 3, 1-31.

Gottlieb, G. (1976). The role of experience in the development of behavior and the nervous system. In G. Gottlieb (Ed.), Neural and behavioral specificity (Vol. 3). New York: Academic Press.

Heinz, J. M., & Stevens, K. N. (1961). On the properties of voiceless fricative consonants. Journal of the Acoustical Society of America, 33, 589-596.

Ingram, D., Christensen, L., Veach, J., & Webster, B. (1980). The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. I, pp. 169-192). New York: Academic Press.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. II, pp. 41-66). New York: Academic Press.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1140.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63, 905-917.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 65, 497-516.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. Perception & Psychophysics, 28, 407-412.

Mann, V. A., Dorman, M. F., Strawhun, D., & Sharlin, H. M. (1982). Development of perceptual adjustment for the coarticulatory effects of rounded vowels on preceding fricatives. Journal of the Acoustical Society of America, 71, S75.

Mann, V. A., & Repp, B. H. (1980). Effect of vocalic context on the [ʃ]-[s] distinction: Perception & Psychophysics, 28, 213-228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, 69, 548-558.

Miller, J. L., & Eimas, P. D. (in press). Biological constraints on the acquisition of language: Further evidence from the categorization of speech by infants. Cognition.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. Perception & Psychophysics, 25, 457-465.

Moskowitz, A. (1975). The acquisition of fricatives: A study in phonetics and phonology. Journal of Phonetics, 3, 141-150.

Prather, E. M., Hedrick, D. L., & Kern, C. A. (1975). Articulation development in children aged two to four years. Journal of Speech and Hearing Disorders, 40, 179-191.

Repp, B. H. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 4, 621-637.

Sevchik, R. A. (1979). An investigation of context dependent effects in the perception of human synthesized speech by Rhesus monkeys (Macca mulatta). Unpublished master's thesis, Department of Psychology, University of Connecticut.

Sharlin, H. M. (1982). The perception of [s] and [ʃ] by children who can and cannot produce these phonemes. Unpublished bachelor's honors thesis, Bryn Mawr College.

Shvachkin, N. K. (1973). The development of phonemic speech perception in early childhood. In C. Ferguson & D. Slobin (Eds.), Studies of child language development (pp. 91-127). New York: Holt, Rinehart, and Winston.

Strange, W., & Broen, P. (1981). The relationship between perception and production of /w/, /r/ and /l/ by three-year-old children. Journal of Experimental Child Psychology, 31, 81-102.

Waters, R. S., & Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. Perception & Psychophysics, 19, 285-289.

TRADING RELATIONS AMONG ACOUSTIC CUES IN SPEECH PERCEPTION: SPEECH-SPECIFIC BUT NOT SPECIAL*

Bruno H. Repp

The perception of most, if not all, phonetic distinctions is sensitive to multiple acoustic cues. That is, there are several distinct aspects of the acoustic speech signal that enable listeners to distinguish between, for example, a voiced and a voiceless stop consonant, or between a fricative and an affricate. Although some cues are more important than others for a given distinction, listeners can usually be shown to be sensitive to even the less important cues when the primary cues are removed or set at ambiguous values. All cues that are relevant to a given phonetic contrast seem to carry information for listeners.

The relevance of a cue can be predicted from comparisons of typical utterances exemplifying the phonetic contrast of interest. Any acoustic property that systematically covaries with a phonetic distinction may be considered a relevant cue for that distinction and may be expected to have a perceptual effect when the conditions are appropriate.

In many recent speech perception experiments several acoustic cue dimensions have been varied simultaneously. Provided the cues are adjusted so that each has an opportunity to influence the perception of the relevant phonetic distinction, it can easily be demonstrated that a little more of one cue can be traded against a little less of another cue, without changing the phonetic percept. This is called a phonetic trading relation.

The perceptual equivalence of acoustically different stimuli obtained by trading two cue dimensions goes beyond the mere equivalence of response distributions. As several recent studies have shown, these stimuli are very difficult to tell apart in a discrimination task. Thus the trade-off among the cue dimensions takes place entirely without the listener's awareness, and only extensive auditory discrimination training might reveal the differences that exist at the auditory level.

Phonetic trading relations are a ubiquitous phenomenon. Whenever two acoustic cues contribute to the same phonetic distinction, they can also be traded against each other, within a certain range. Thus, these trading relations are a manifestation of a more general perceptual principle of cue integration, by which I mean the assumption that, in phonetic perception, the

---

*This position paper was presented at the Tenth International Congress of Phonetic Sciences in Utrecht, August 1983.

Acknowledgment. Preparation of this paper was supported by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

information conveyed by a variety of acoustic cues is integrated and combined into a unitary perceptual experience that can be described in terms of linguistic categories. But what causes the various cues to be integrated and to trade with each other?

One current school of thought holds that the integration of cues and the ensuing trading relations are due to auditory interactions of one sort or another. Proponents of this hypothesis, while ready to admit that the psychoacoustics of complex speech signals are not yet well understood, nevertheless believe that phenomena known from research with nonspeech stimuli, such as auditory adaptation, masking, or integration, can account for trading relations in speech. Opponents of this hypothesis, on the other hand, like to point out the great acoustic diversity of the cues involved and their distribution over considerable temporal intervals. Obviously, and especially as far as specific trading relations are concerned, this dispute can only be settled by empirical research. A number of recent experiments have addressed this issue, employing several different techniques, but which are alluded to in my abstract. I do not have the time to summarize the results here; suffice it to say that the available evidence suggests that many phonetic trading relations occur only when listeners engage in the phonetic classification of speech signals, and not when they identify analogous nonspeech stimuli or discriminate auditory properties of speech. Thus these trading relations seem to be a product of phonetic categorization, not of interactions in the auditory system. This is not to say that auditory interactions do not occur in speech signals, although it is possible that, due to the intimate familiarity of listeners with speech, such interactions have less of a perceptual impact than in less familiar nonspeech stimuli. Certain effects of irrelevant signal properties on phonetic perception do seem to require a psychoacoustic explanation. And indeed, some of the many trading relations that now appear to be phonetic in origin may eventually be proven to rest on an auditory interaction. It seems extremely unlikely, however, that all of them will be so explained.

The reason for this prediction of mine is that psychoacoustic approaches to speech perception often seem to ignore a crucial fact--that phonetic classification takes place with reference to norms established through past experience with a language. Although this experience has been filtered and transformed by the constraints and nonlinearities in the auditory system through which it had to pass, the current input undergoes precisely the same transformations, so that the topological relationship between it and the internal representation of past experience remains essentially unchanged. It is this relationship that determines the phonetic percept by a principle of proximity: The input is perceived as whatever it resembles most in the past experience of the individual. There is, of course, much more to be learned about the perceptual metric that relates speech stimuli and the representations of phonetic categories in the listener's mind, and auditory nonlinearities may indeed influence that metric. The essential point, however, is that the perception of phonetic categories derives from a relationship, and not from any properties of the acoustic signal per se. Neither the relevance nor the perceptual importance of acoustic cues can be predicted from an inspection of the input alone. Rather, the integration and weighting of the cues is a perceptual function based on a relationship of input to knowledge within the speech domain. Phonetic trading relations are, therefore, necessarily a speech-specific phenomenon, even if certain individual trading relations could potentially (or do in fact) arise from auditory interactions. As we learn

more about the peripheral auditory transformations of speech signals, we may eventually be able to redefine the perceptual cues in a way that makes the trading relations among them exclusively phonetic.

Having argued for the speech-specificity of phonetic trading relations, I would now like to address the question of whether the perceptual integration of phonetically relevant cues is achieved by some special machinery or process, or whether it reflects a general principle of perception. In the past, it has often been argued that speech perception makes reference to speech production, and that perceptual processes actually make use of some of the neural networks engaged in articulation. This certainly remains an interesting and important hypothesis at the neurophysiological level. To the perceptual theorist, however, it really should be a truism: Since speech perception occurs with reference to internal criteria based on language experience, and since language is produced in a systematic manner by human vocal tracts, the listener's internal representation of past experience with his or her language necessarily embodies articulatory constraints as well as language-specific characteristics. In other words, I would like to argue that speech perception must reflect the way speech is produced because the criteria for perceptual classification are the production norms of the language. To say, therefore, that speech perception refers to speech production is merely to state the obvious.

A more specific hypothesis regarding phonetic trading relations might be proposed, however. It might be argued that many individual cues that trade in perception also trade in production, in the sense that there is a continuous covariation of the two acoustic cues, due to some articulatory reciprocity, even within phonetic categories. If it were the case that perceptual trading relations are obtained only for cues that show such continuous covariation in production, then it might be argued that speech perception makes use of specific knowledge of patterns of articulatory variability, and since the brain presumably cannot store an infinity of variants, it might be inferred that reference is made to an internal representation of the articulatory mechanism that enables listeners to generate specific cue relationships. Although this hypothesis needs to be explored in greater depth, it seems to me that the continuous covariation of cues in production should not be a necessary condition for perceptual trading relations to occur. All that is required is that typical instances of two different phonetic categories differ along two or more acoustic dimensions. It is much more plausible and parsimonious to assume that the listener's brain retains a record of typical instances of utterances, that is of the central tendencies in the variability encountered, rather than of the variability itself. While this system of phonetic category prototypes must be adjustable to the changing characteristics of ongoing speech, at any given point in time it provides the stable reference points that guide speech perception.

From this broad vantage point, phonetic classification is a form of pattern recognition. Speech signals may be thought of as points or traces in a multidimensional auditory space that also harbors the appropriately tuned category prototypes, and phonetic categories are selected on the basis of some distance metric. Trading relations among the various acoustic dimensions of this auditory-phonetic space are an obvious consequence. What makes speech special, in this view, is not the processes or mechanisms employed in its

perception but the unique structure of the patterns that are to be recognized, which reflect in turn the special properties of the production apparatus and the language-specific conventions according to which it is operated.

In summary, then, I have argued that, on the one hand, phonetic trading relations are speech-specific but, on the other hand, they are not special as a phenomenon. They are speech-specific because their specific form can only be understood by examining the typical patterns of a language. They are not special because, once the prototypical patterns are known in any perceptual domain, trading relations among the stimulus dimensions follow as the inevitable product of a general pattern matching operation. Thus, speech perception is the application of general perceptual principles to very special patterns.

THE ROLE OF RELEASE BURSTS IN THE PERCEPTION OF [s]-STOP CLUSTERS*

Bruno H. Repp

Abstract. The role of the release burst as a cue to the perception of stop consonants following [s] was investigated in a series of studies. Experiment 1 demonstrated that silent closure duration and burst duration can be traded as cues for the "say"-"stay" distinction. Experiment 2 revealed a similar trading relation between closure duration and burst amplitude. Experiments 3 and 4 suggested, perhaps surprisingly, that absolute, not relative, burst amplitude is important. Experiment 5 demonstrated that listeners' sensitivity to bursts in a labeling task is at least equal to their sensitivity in a burst detection task. Experiments 6 and 7 replicated the trading relation between closure duration and burst amplitude for labial stops in the "slit"-"split" and "slash"-"splash" distinctions, although burst amplification, in contrast to attenuation, had no effect. All experiments revealed that listeners are remarkably sensitive to the presence of even very weak release bursts.

Introduction

A large proportion of speech perception research has been concerned with stop consonants. Nevertheless, there are still gaps in our knowledge of the relevant acoustic cues and their perceptual importance. While much attention has been lavished on the perception of stop consonant voicing and place of articulation, the more basic question of whether or not a stop consonant is perceived at all has been addressed in only a handful of studies. Moreover, nearly all of these studies have used synthetic speech stimuli in which at least one important cue was commonly absent: the release burst that terminates the stop closure. The present series of studies explores the role of this cue in the perception of stop consonants after [s].

A good deal is known about some other cues to stop manner perception, at least in the context of preceding [s] and following vowel or [l]. One very important cue is an interval of silence corresponding to the period of oral closure that characterizes stop consonant articulation. Early research at Haskins Laboratories by Bastian (1959, 1962) as well as the recent thorough investigations of Bailey and Summerfield (1980) have shown that an interval of

---

*Also Journal of the Acoustical Society of America, in press.
Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR05596 to Haskins Laboratories. Results of Experiments 1 and 2 were reported at a meeting of the Acoustical Society of America in Orlando, FL, November 1982. I am grateful to Joanne Miller, Ralph Ohde, Sigfrid Soli, and Douglas Whalen for helpful comments on an earlier draft.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

silence between an [s]-noise and a steady-state synthetic vowel is generally sufficient to elicit a stop consonant percept, given that the silence is longer than about 20 ms (but not excessively long), and that the vowel is not too open. Silence frequently is not only sufficient but also necessary for the perception of a stop, for even when other stop manner cues are present in the signal (neglecting release bursts for the moment), stops are rarely perceived in the absence of an appropriate closure interval (Bailey & Summerfield, 1980; Best, Morrongiello, & Robson, 1981; Dorman, Raphael, & Liberman, 1979; Fitch, Halwes, Erickson, & Liberman, 1980).

Other relevant cues reside in the signal portions adjacent to the closure interval. Changes in spectrum and/or a rapid amplitude drop in the preceding fricative noise signify the approach of the closure and thereby contribute to stop perception (Repp, unpublished data; Summerfield, Bailey, Seton, & Dorman, 1981). Similarly, formant transitions and/or a rapid amplitude rise at the onset of the following vocalic portion—a rising transition of the first formant (F1) in particular—signify rapid opening and thereby constitute an important stop manner cue (Bailey & Summerfield, 1980; Best et al., 1981; Fitch et al., 1980). There is also evidence that the durations of the acoustic segments preceding and following the closure can influence stop manner perception (Summerfield et al., 1981; however, see also Marcus, 1978). These additional cues engage in trading relations with the temporal cue of closure duration; that is, the stronger they are, the less closure silence is needed to perceive a stop. (For analogous findings for stops in vowel-[s] context, see Dorman, Raphael, & Isenberg, 1980.) In general, however, these studies suggest that a minimal amount of silence (about 20 ms) is needed for a stop to be perceived at all.

Nearly all of the above-mentioned studies used synthetic speech stimuli that did not include any release bursts. One reason for this omission was presumably that good bursts are difficult to synthesize. Although most researchers are probably aware of the relevance of release bursts to the perception of stop manner, the importance of this cue has not been sufficiently acknowledged in the literature, which has emphasized the role of the closure duration cue. In an unpublished study, Repp and Mann (1980) took three tokens each of [sta], [ska], [ʃta], and [ʃka], produced by a male speaker, excised the closure period, and replaced the natural fricative noises with synthetic ones of comparable amplitude. In one condition, the stimuli retained the natural release bursts, and the subjects continued to report stop consonants on 100 percent of the trials, with very few place-of-articulation errors. In another condition, the release bursts were excised, and stop responses fell to 3 percent (except for two subjects who continued to report stops, but with poor accuracy for place of articulation). These data clearly illustrate the salience of the release burst as a manner cue for alveolar and velar stops following fricatives. Labial stops, on the other hand, are associated with weaker release bursts (see Zue, 1976) that may not be sufficient to cue a stop percept in the absence of an appropriate closure interval.

The present series of studies attempts to answer several questions about the role of release bursts in stop manner perception: (1) Given that an interval of silence is needed to hear an alveolar stop when there is no release burst but not when there is one, how much can the burst cue be weakened before any silence is needed, and will further weakening of the burst result in increasing amounts of silence required? In other words, how sensitive are listeners to burst cues, and is there a regular trading relation between the

burst and silence cues? These questions are explored in Experiments 1 and 2 by manipulating alveolar burst duration and amplitude. (2) Given an effect of burst amplitude that can be traded against silence duration, Experiments 3 and 4 investigate whether it is absolute or relative burst amplitude that matters. (3) Experiment 5 addresses the question of whether the point at which an attenuated release burst ceases to trade with silence coincides with the auditory detection threshold for the burst. (4) The role of burst amplitude is further investigated in Experiments 6 and 7 with labial stops, with special attention to the question of whether amplification of a weak labial burst can make it a more powerful manner cue.

Experiment 1

The purpose of Experiment 1 was to demonstrate the relative importance of an alveolar release burst as a stop manner cue, and to create a trading relation between burst and silence cues by varying the durations of both in natural-speech stimuli.

Method

Stimuli. Good tokens of "say" and "stay" were selected from recordings of several repetitions produced by a female speaker in a sound-insulated booth. These two utterances were low-pass filtered (-3 dB at 9.6 kHz, -55 dB at 10 kHz), digitized at 20 kHz, and modified by waveform editing. To reduce stop manner cues in the fricative noise, which were not of particular interest in the present study, the [s]-noise of "say," 176 ms in duration, was used in all experimental stimuli. This noise was followed by a variable interval of silence and by one of seven different, "day"-like portions, roughly 550 ms in duration. Six of these were derived from the token of "stay" while the seventh represented the vocalic portion of the "say" token.

Figure 1 shows the waveforms of the onsets of these stimulus portions. On top is the original post-closure portion of "stay," which began with a rather powerful (but, for that speaker, not atypical) release burst of somewhat less than 20 ms in duration. The rms amplitude of the total burst was determined to be 4.6 dB below the vowel onset and 6.8 dB below the vowel peak (135 ms later), with an amplitude decrease of about 10 dB from the initial to the final quartile of the burst.[2] The release burst was cut back in five steps, as indicated in the figure. Successive cuts (versions 2-6) were made at 6.1, 10.6, 13.4, 15.2, and 19.6 ms from the onset. These cutpoints were selected visually on the basis of local dips in the waveform. In each case, the cut was made at the nearest zero crossing. The stimulus portion derived from "say" is shown at the bottom of Figure 1, aligned so as to show its similarity with version 5 of the "day" portion on top. Despite this similarity of waveforms, however, there were presumably some spectral differences between these two portions, due to the different contexts in which they had been articulated.

The silent interval separating the initial fricative noise from the "day" portions was varied from 0 to 60 ms in 10-ms steps. Because tokens with large bursts were expected to be perceived as "stay" even without any silence, a semi-orthogonal design was employed that assigned an increasingly wider range of silence durations to tokens with increasingly shorter bursts. Thus the stimuli with the most powerful burst occurred only with the 0-ms silence, while the stimulus derived from "say" occurred with all seven silent intervals. This

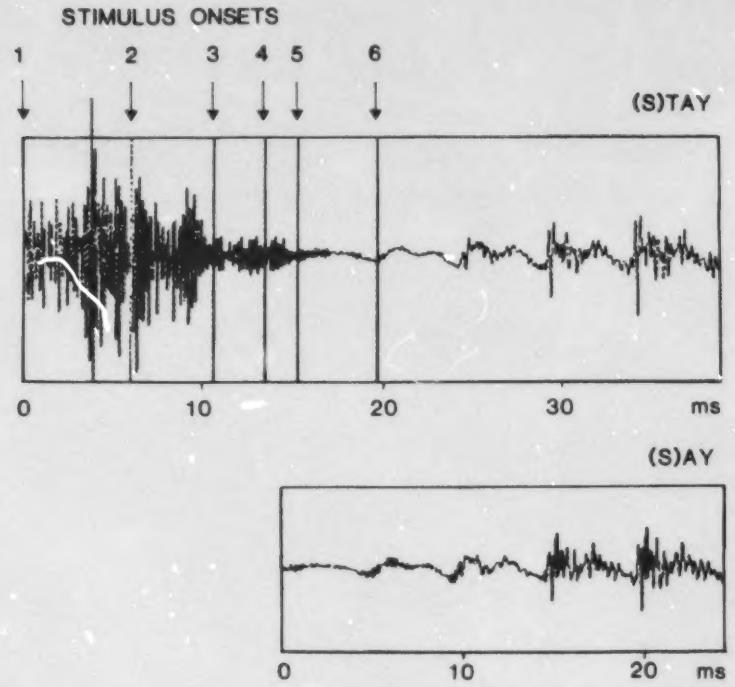


Figure 1. Onset waveforms of stimuli used in Experiment 1. The top panel shows the first 40 ms following the closure in "stay"; the bottom panel shows the first 24 ms following the fricative noise in "say". Arrows in the top panel indicate cut points for release burst truncation.

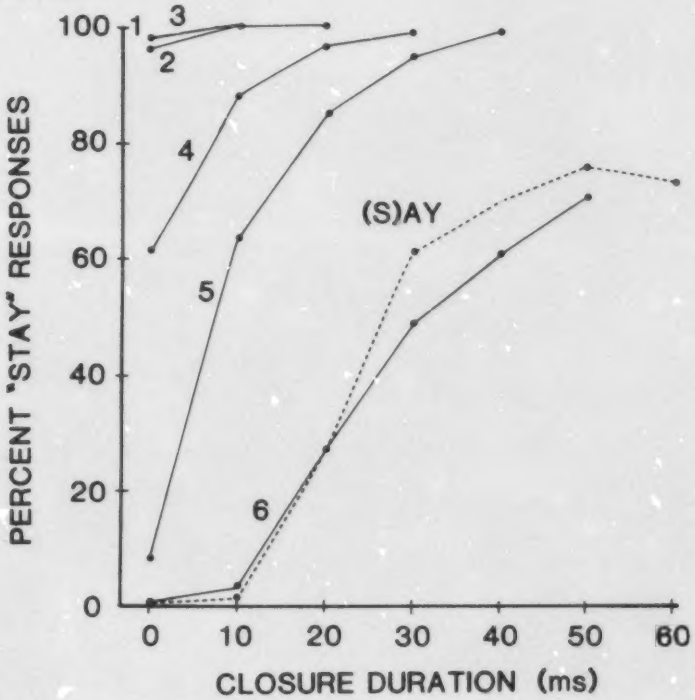


Figure 2. Trading relation between alveolar release burst duration and closure duration (Exp. 1). Numbers refer to cut points illustrated in Figure 1. The dashed line represents the token derived from "say". Closure duration (abscissa) refers to the actual silence in the stimuli.

led to a total of 28 different stimuli that were recorded on audio tape in 10 different randomizations, with interstimulus intervals of 2 s.

Subjects and procedure. Ten subjects participated, including nine paid volunteers and one member of the laboratory staff (not a speech researcher). None of the subjects reported any hearing problems, and all had only very limited experience in speech perception experiments. The stimuli were presented binaurally over calibrated TDH-39 earphones in a quiet room.[3] The subjects identified in writing each stimulus as either "say" or "stay."

Results and Discussion

Average percentages of "stay" responses are shown as a function of silent closure duration in Figure 2, separately for each of the seven stimulus patterns. It is evident that versions 1, 2, and 3 were invariably identified as "stay," even in the absence of silence. Thus, even the remainder of the burst following the initial high-amplitude portion (version 3, see Figure 1) was a sufficient cue for stop manner. As the burst was cut back further, increasing amounts of silence were necessary to achieve a percept of "stay." The stimulus with the "say"-derived portion yielded results similar to those for version 6, and it appears that neither provided sufficient cues for unambiguous "stay" percepts, even at the longest silences used here.

What is most striking about these results is the large perceptual effect that a small burst cutback had on perception. The change from version 4 to version 5 consisted of the elimination of only 1.8 ms of relatively low-amplitude noise at onset (see Figure 1); however, listeners needed approximately 10 ms more silence to compensate for this loss and achieve the same average rate of "stay" responses. Similarly, the change from version 5 to version 6 consisted of the elimination of the last 4.4 ms of burst residue. The perceptual effects were dramatic: At least 20 ms of additional silence were needed to compensate for the loss, and several listeners were not able to compensate for it at all, reporting only "say" for version 6. Even those few subjects who did reach a 100-percent "stay" asymptote for version 6 and had very steep labeling functions showed large effects of the stimulus manipulations.

Thus, this study not only demonstrates a perceptual trading relation between burst duration and silence duration but also that listeners are remarkably sensitive to what seem to be rather minute changes in the onset characteristics of the stimulus portion following the silent closure interval. Of course, the truncation of the release burst introduced not only variations in burst duration but also changes in overall burst amplitude, in its onset amplitude characteristics, and perhaps correlated spectral changes. Any of these may have been responsible for the effects observed, but it is still true that relatively small physical changes had relatively large perceptual consequences.

Experiment 2

Experiment 2 examined one parameter that may have played a role in Experiment 1--the overall burst amplitude. The purpose of the study was to demonstrate a trading relation between release burst amplitude and closure duration as joint cues to stop manner perception.

Method

Stimuli. In Experiment 1, stimulus version 3 was just on the verge of requiring some silence in addition to the truncated burst, in order for a stop to be perceived on all trials (see Figure 2). This stimulus was chosen as the starting point. Its residual burst was 9 ms in duration (see Figure 1), with a total rms amplitude 10.8 dB below the vowel onset and 15.1 dB below the vowel peak. Five additional versions were created by digitally attenuating the burst by up to 30 dB in 6-dB steps. In a seventh version the burst was infinitely attenuated (i.e., it was replaced with 9 ms of silence); thus this stimulus was equivalent to stimulus version 6 in Experiment 1.

Silent intervals ranging from 0 to 60 ms in 10-ms steps were assigned to the stimuli using the same design as in Experiment 1. Thus, version 1 occurred only with the 0-ms interval while version 7 occurred with the full range of closure durations. The resulting 28 stimuli were recorded in 10 different randomizations.

Subjects and procedure. Twelve new subjects participated in this study. The data of one had to be discarded because he could not reliably distinguish among the stimuli. The remaining eleven subjects included eight staff members of Haskins Laboratories (including the author) with varying amounts of experience in speech perception tasks, and three paid student volunteers. The procedure was the same as in Experiment 1.

Results and Discussion

Average percentages of "stay" responses are shown as a function of silent closure duration in Figure 3, separately for each of the seven attenuation conditions. It is evident that there is an orderly progression of labeling functions: As the burst got weaker, more silence was needed to perceive a stop consonant.

The figure suggests that a burst attenuated by as much as 30 dB still led to more stop responses than a stimulus without any burst. This was confirmed in a one-way analysis of variance on the stop responses to these two types of stimuli, summed over closure durations of up to 40 ms, $F(1,10) = 9.8$, $p < .02$. Since, in the 30-dB attenuation condition, the amplitude of the 9-ms residual burst was about 45 dB below the vowel peak amplitude (or at about 38 dB SPL versus about 83 dB SPL for the vowel at the subjects' earphones), this finding again reveals that listeners are remarkably sensitive to burst cues.

Two additional comments are in order concerning Figure 3. First, it should be noted that, in the infinite attenuation condition, the nominal closure ended at the beginning of the nonexistent burst. Therefore, the actual duration of the silence in these stimuli was 9 ms longer, as indicated by the arrows in the figure, which makes the results more nearly comparable to those for the same stimulus (version 6) in Experiment 1 (see Figure 1). It would not have been appropriate to plot these data in terms of actual silence duration because the effective silence durations resulting from various degrees of burst attenuation are not known. Note, however, that such a plot would tend to space the functions in Figure 3 farther apart and thus increase the observed effects. This distinction between nominal and actual closure duration will recur in later experiments.

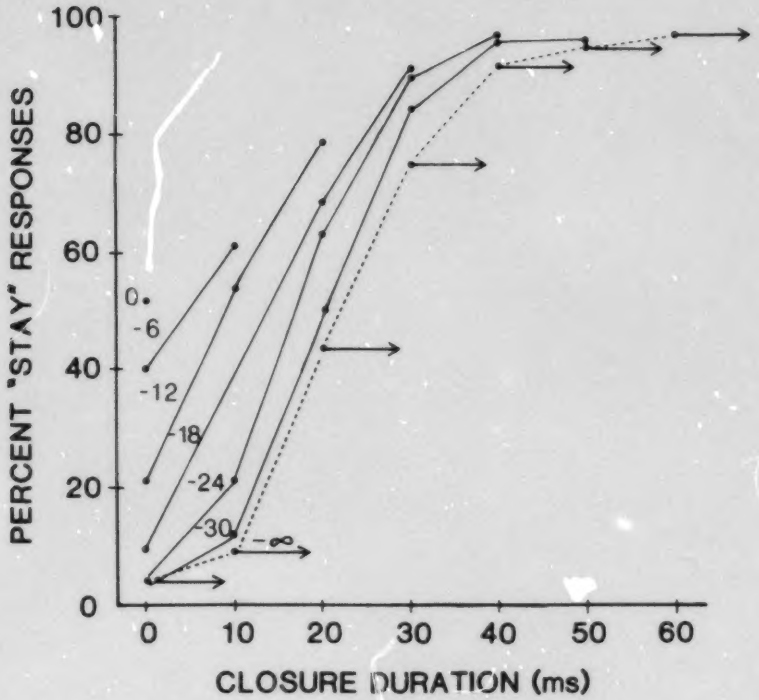


Figure 3. Trading relation between alveolar release burst amplitude and closure duration (Exp. 2). Negative numbers refer to amplitude decrement (in dB). Closure duration (abscissa) is nominal; the actual silence durations in the infinite-attenuation condition were 9 ms longer due to the silenced burst, as indicated by the arrows.

Second, it will be noted that, contrary to expectations based on Experiment 1, the unattenuated stimulus did not receive 100 percent "stay" responses, while the burstless stimulus did reach this asymptote at the longer silences. Although there was considerable variability among individual subjects with regard to how the unattenuated stimulus was perceived, the pattern of the data suggests that the subjects gave somewhat more weight to closure duration and less weight to the burst in Experiment 2 than in Experiment 1. The reason for this is not known.

In summary, the present study demonstrated the expected trading relation between burst amplitude and closure duration, and it showed that severely attenuated (and truncated) bursts still can have a perceptual effect.

Experiment 3

Given the finding of the preceding study that burst amplitude is an important parameter, Experiment 3 addressed the question of whether the perceptually relevant aspect of burst amplitude is its absolute magnitude or its magnitude relative to the surrounding signal portions.

Method

Stimuli. Taking the data of Experiment 2 as a guideline, the stimulus with the 12-dB attenuation of the 9-ms residual burst was selected as the starting point for the present study. Four other stimuli were created by

selectively attenuating portions of this original stimulus, as illustrated schematically in the upper right-hand corner of Figure 4. In addition to (a) the original stimulus, there were stimuli with attenuation of (b) only the burst, (c) both the burst and the following vocalic portion, (d) the burst and the preceding fricative noise, and (e) the whole stimulus. Attenuation was by 12 dB in all cases.

All stimuli occurred with all closure durations, which varied from 0 to 40 ms in 10-ms steps. The resulting 25 stimuli were recorded in 10 different randomizations.

Subjects and procedure. Ten subjects participated, including six new paid volunteers and four staff members of Haskins Laboratories (including the author). Results were similar for the two groups of subjects and were combined. One subject reported only "say" during the first half of the test, so only her data from the second half were included. The procedure was the same as in Experiments 1 and 2.

Results and Discussion

The labeling functions for the five conditions are drawn in the top panel of Figure 4. Clearly, the stimulus manipulations made a difference. This was confirmed by a one-way analysis of variance on the percentages of "stay" responses summed over all closure durations, $F(4,36) = 12.6$, $p < .001$. Statistical comparisons among individual conditions were done by post-hoc Newman-Keuls tests. According to these tests, condition (a) differed significantly ($p < .01$) from all other conditions, and condition (c) differed ($p < .05$) from condition (d).

A graphic comparison among conditions is provided in the bottom part of Figure 4 in terms of the location of the average "say"-"stay" boundary (obtained by linear interpolation between the data points straddling the boundary) on the closure duration dimension. Proceeding from left to right through the five panels, we see the following: (1) Attenuation of the fricative noise, holding the other stimulus components constant, increased the number of stop responses slightly (i.e., the boundary shifted to a shorter silence duration). (2) Attenuation of the burst decreased stop responses substantially, which replicates Experiment 2. (3) Attenuation of the voiced portion resulted in a slight decrease in stop responses. (4) Attenuating both the fricative noise and the voiced portion together had absolutely no effect. (5) Attenuation of the whole stimulus caused a substantial decrease in stop responses equivalent to that resulting from attenuation of the burst alone.

These results point toward absolute burst amplitude as the relevant factor. Clearly, attenuating the burst's environment did not have the same effect as amplifying (more precisely, restoring) the burst by the same amount (see Figure 3). Contrary to expectations, attenuation of the vocalic portion did not increase stop responses. Perhaps, additional stop manner cues contained in that portion (initial formant transitions and amplitude envelope) were weakened by the attenuation, thus counteracting the gain in burst salience relative to its environment. If so, however, we are forced to conclude that the absolute amplitude of those cues matters, which is equally interesting.

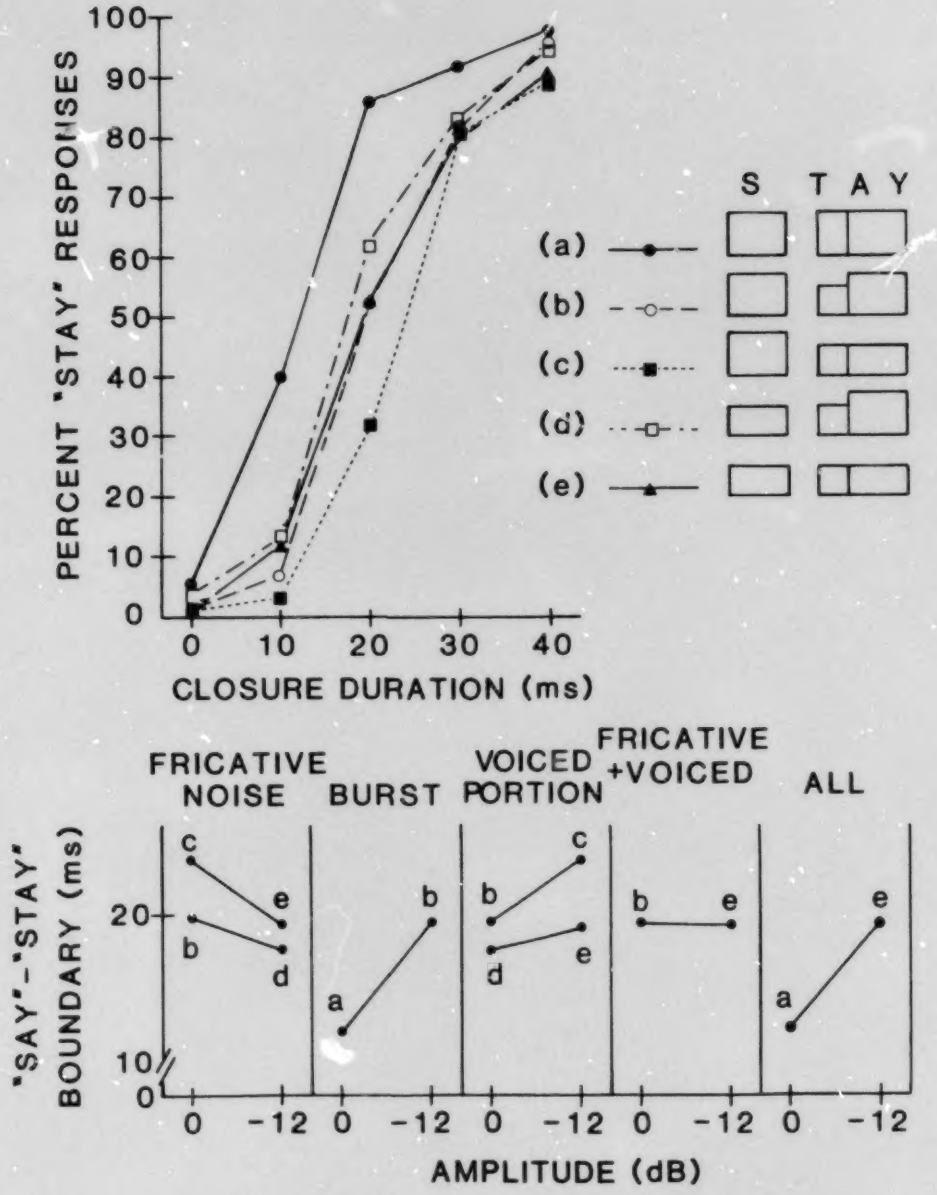


Figure 4. Design and results of Experiment 3. Labeling functions for the five conditions are provided on the upper left, with the key on the upper right. Rectangles in the key represent schematically the "s" fricative noise, the "t" burst, and the "day" voiced portion. The height of the rectangles represents amplitude relative to the base stimulus (condition a). At the bottom, comparisons among the various conditions are presented in terms of average category boundary values (in ms of closure silence). Lower-case letters refer to the key on top.

Another possibility is that the present study suffered from floor effects due to listeners' inability to detect the burst when it was attenuated. This would explain why the largest difference occurred between condition (a) and all others. Note that conditions (a) and (b) were equivalent to the 12-dB and 24-dB attenuation conditions in Experiment 2. The average category boundaries for these conditions were at 9 and 17 ms, respectively, in Experiment 2, and at 10 and 20 ms in Experiment 3--a rather close agreement. Note also that, in Experiment 2, a burst attenuated by 24 dB still had a significant perceptual effect. The agreement between Experiments 2 and 3 suggests that the absolute stimulus amplitudes were similar, and that no floor effect occurred. Nevertheless, it seemed advisable to replicate the present results with the burst amplitude set at somewhat higher absolute levels, and with inclusion of a no-burst baseline condition.

Experiment 4

This replication of Experiment 3 used new stimuli in a complete 2 x 3 x 2 orthogonal design. By including burstless stimuli in the design, it was possible to examine the effects of fricative noise and vowel attenuation separate from their effects on the relative salience of the burst--an important control condition.

Method

Stimuli. Good tokens of "say" and "stay" were selected from among several repetitions recorded by a new female speaker. Both utterances were digitized at 20 kHz. As in Experiments 1-3, the fricative noise (170 ms long) was taken from "say." The "day" portion of "stay" was about 450 ms in duration and began with a release burst 13.35 ms long. The overall rms amplitude of this burst was determined to be 5.5 dB below the vowel onset, 11.0 dB below the vowel peak (only 20 ms later), and 4.1 dB above the fricative noise maximum. Informal listening confirmed that this burst, as usual, was sufficient for "stay" to be perceived without any closure silence (see also Exp. 5). To be able to trade burst amplitude against silence, the most intense burst used was 15 dB below the original. A total of 12 stimulus versions were created by orthogonally combining three factors: fricative noise attenuation (0 or 10 dB), burst attenuation (15 or 25 dB, or no burst at all),[4] and "vowel" attenuation (0 or 10 dB). Each of these 12 versions occurred with five closure durations ranging from 0 to 40 ms in 10-ms steps. The resulting 60 stimuli were recorded in 5 different randomizations.

Subjects and procedure. Ten new paid volunteers identified the stimuli as "say," "stay," "spay," or "svay." The last two response alternatives were included because the author, as a pilot subject, had noticed a tendency to hear these additional categories. The tape was repeated once, so that each subject gave ten responses to each stimulus.

Results and Discussion

Of the ten subjects, three gave only "say" and "stay" responses, while the other seven used one or both of the additional response categories as well. In the initial analysis, all consonant cluster responses were pooled.

Since the burstless stimuli had been created by omitting the burst rather than by infinitely attenuating it, 13.35 ms (the duration of the burst) must be subtracted from their actual closure durations to compare results directly for stimuli with and without bursts. This has been done graphically in Figure 5, where the arrows point toward the actual closure durations. The figure shows average labeling functions for the three burst conditions, averaged over fricative and vowel attenuation conditions. Clearly, the subjects gave many more cluster responses to the stimuli with bursts than to those without. Elimination of the burst resulted in a flattening of the labeling function; 40 ms of silence was not enough to make a burstless stimulus sound like an unambiguous "stay." The figure also shows the expected effect of the 10-dB burst attenuation. It is clear that this experiment avoided the danger of floor effects; if anything, the burst amplitudes were somewhat too high.

The effects of variations in fricative noise and vowel amplitude, which were smaller than the effects of burst amplitude, are summarized in Table 1 in the form of response percentages averaged over all closure durations. A three-way repeated-measures analysis of variance (with the factors Burst, Fricative, and Vowel) was first conducted on the total cluster responses, ignoring the incommensurability of actual closure durations for stimuli with and without bursts. This analysis revealed, besides the expected Burst effect, $F(2,18) = 64.8$, $p < .001$, significant main effects of both Fricative, $F(1,9) = 9.6$, $p < .02$, and Vowel, $F(1,9) = 12.4$, $p < .01$, as well as significant interactions between Burst and Fricative, $F(2,18) = 10.7$, $p < .001$, and between all three factors, $F(2,18) = 12.1$, $p < .001$. To clarify the triple interaction, separate analyses of variance were conducted on stimuli with and without bursts. Stimuli with bursts exhibited significant main effects of Burst, $F(1,9) = 32.7$, $p < .001$, Fricative, $F(1,9) = 45.2$, $p < .001$, and Vowel, $F(1,9) = 10.4$, $p = .01$, as well as a marginal Burst by Fricative interaction, $F(1,9) = 5.4$, $p < .05$, and a strong triple interaction, $F(1,9) = 30.6$, $p < .001$. Thus, the triple interaction was not due to different patterns of results for stimuli with and without bursts. The separate analysis of burstless stimuli revealed only a significant effect of Vowel, $F(1,9) = 8.5$, $p < .02$, not of Fricative.

Consider now the directions of these effects. The Burst effect, of course, was due to a decrease of cluster responses as the burst was attenuated or eliminated altogether (Figure 5). The Fricative effect, too, was in the expected direction: Attenuation of the fricative noise <u>increased</u> the number of cluster responses. (A similar but nonsignificant trend was observed in Exp. 3.) This is the kind of effect that might be expected if the fricative noise reduced the salience of the burst through some form of auditory forward masking (see Delgutte, 1980). This interpretation is supported by the finding that the Fricative effect was absent in burstless stimuli, where there was no burst to be masked (see Table 1).

Turning now to the Vowel effect, it can be seen in Table 1 that attenuation of the vocalic portion, like attenuation of the fricative noise, resulted in an <u>increase</u> of cluster responses, contrary to a nonsignificant opposite trend observed in Experiment 3. Since this was true regardless of whether a burst was present or absent, the effect was apparently not due to release from a backward masking effect of the vowel on the release burst, or simply to an increase in the salience of the burst relative to the vowel.[5]

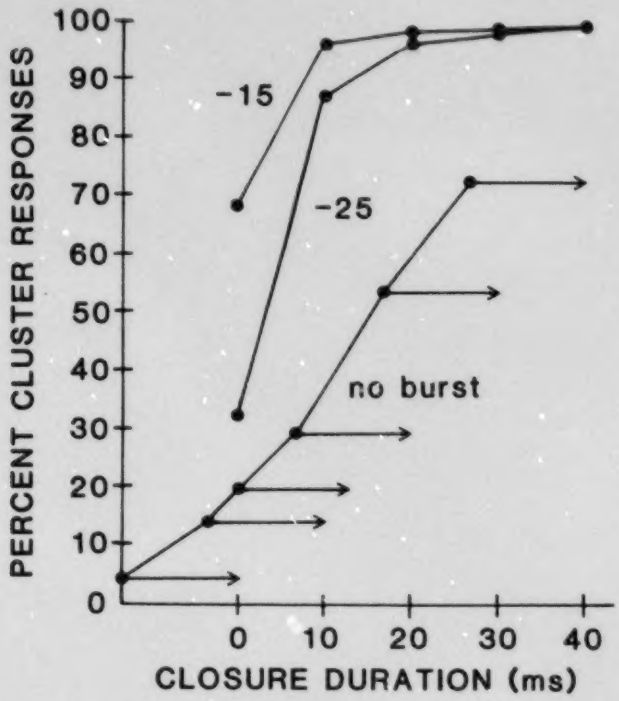


Figure 5. Effect of burst amplitude in Experiment 4, averaged over other amplitude conditions. Numbers refer to burst amplitude (in dB) relative to original burst. Closure durations are nominal; actual silence durations in the no-burst condition are indicated by arrows.

Table 1

Response pattern in Experiment 4, averaged over closure durations.

Stimulus amplitude (dB)			Response (percent)				
Burst	Fricative	Vowel	"say"	"stay"	"svay"	"spay"	Total cluster
-15	0	0	16.6	79.8	3.0	0.6	83.4
	-10	0	6.8	87.4	3.4	2.4	93.2
	0	-10	8.2	90.8	1.0	0.0	91.8
	-10	-10	1.2	97.4	0.8	0.6	98.8
-25	0	0	20.2	57.4	16.8	5.6	79.8
	-10	0	19.6	57.2	17.8	5.4	80.4
	0	-10	20.2	74.8	3.6	1.4	79.8
	-10	-10	10.4	86.0	2.2	1.4	89.6
no burst	0	0	67.6	12.4	10.0	10.0	32.4
	-10	0	70.8	7.0	11.4	10.8	29.2
	0	-10	60.8	25.4	5.2	8.6	39.2
	-10	-10	61.8	20.8	8.8	8.6	38.2

The results are complicated by the triple interaction, which was due to the fact that, with the higher burst amplitude, fricative and vowel attenuation seemed to have independent effects whereas, with the lower burst amplitude, only simultaneous attenuation of both produced an effect. An explanation of this complex pattern is beyond reach at the moment.[6]

In summary, this experiment, in conjunction with Experiment 3, provides little support for a role of relative burst amplitude in stop manner perception. While the preceding fricative noise may exert a slight masking effect on the burst, the amplitude of the following vocalic portion seems to have its perceptual effects primarily by changing the relative salience of cues contained in that portion itself. While the present data cannot be considered the last word on the issue, the possibility of a fixed perceptual criterion in the amplitude domain deserves further attention, both with regard to the perception of stop manner and to place-of-articulation distinctions in stops (see Ohde & Stevens, 1983) and fricatives (Gurlekian, 1981).

Experiment 5

The preceding experiments, Experiment 2 in particular, demonstrate a remarkable sensitivity of listeners to the presence of even very weak release bursts. This suggests the hypothesis that the point at which a burst becomes ineffective and ceases to trade with closure silence actually coincides with the auditory detection threshold for the burst. This hypothesis was tested in the present experiment.[7] In addition, the study examined whether the detectability of the burst is increased when the preceding fricative noise is removed.

Method

Stimuli. The stimuli were derived from the utterances that also provided the basis for the stimuli of Experiment 4. In addition to the original stimulus (full burst amplitude), six levels of burst attenuation were employed: 10, 20, 25, 30, 35, and ∞ dB. In the identification test, these seven stimuli occurred with nominal closure durations of 0, 10, 20, and 30 ms. Ten different randomizations of the 28 stimuli were recorded.

In addition, two discrimination tests were assembled, which required subjects to detect the presence of a burst. The two tests were identical except that in one the initial fricative noise was omitted from all stimuli while, in the other, the fricative noise was followed by a fixed 10-ms closure interval. A fixed-standard same-different paradigm was employed. The fixed standard was the burstless stimulus; it occurred first in every stimulus pair. After a fixed interval of 500 ms, the comparison stimulus occurred; it either did or did not contain a release burst. Over six successive test blocks, the burst in the comparison stimulus was attenuated by 0, 10, 20, 25, 30, and 35 dB. Each test block consisted of 50 trials, the first 10 of which were practice, with the responses alternating between "same" and "different" and known in advance. Half of the remaining 40 trials were "same" and half were "different," in random order. The intertrial interval was 2 s.

Subjects and procedure. Ten paid volunteers participated in the experiment, six of whom had also been subjects in Experiment 4. In the identification test, which was always presented first, they responded "say" or "stay," with "svay" and "spay" as additional options. In the discrimination tests,

the responses were "s" ("same") and "d" ("different"). The order of the two discrimination tests was counterbalanced across subjects. Playback amplitude was controlled by adjusting the level so as to achieve a constant maximum deflection on a vacuum tube voltmeter, and by keeping it at that level throughout the experiment. All tapes had been recorded at the same level. The peak amplitude of the vowel (and, hence, of the unattenuated burst as well—see Exp. 4) at the subjects' earphones was estimated to be approximately 83 dB SPL.

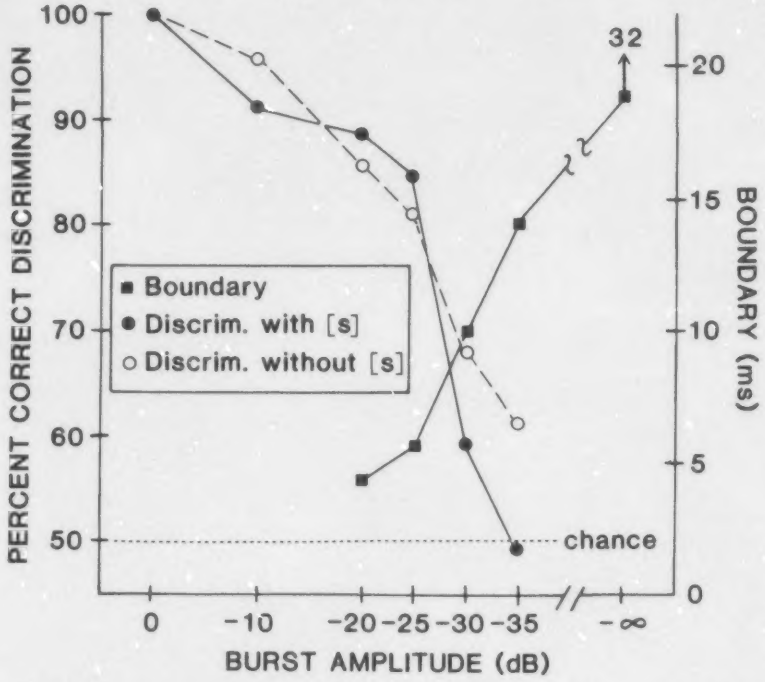


Figure 6. Identification and burst detection results from Experiment 5. Filled squares plot the category boundary in ms of silence (right ordinate) as a function of burst amplitude. Silence duration for the no-burst condition is nominal; actual silence at the boundary was 32 ms, as indicated. Circles show burst detection scores as percent correct (left ordinate) for stimuli with and without initial [s]-noise.

Results and Discussion

The average data are presented in Figure 6. The labeling boundary for stimuli with bursts attenuated by 20 dB or more is represented by the filled squares. Stimuli with an unattenuated burst were uniformly identified as "stay," and those with a -10 dB burst received only 16 percent "say" responses when no silence was present, so no boundaries could be determined for these stimuli. As expected, the boundary shifted toward increasingly longer values of silence as the burst was attenuated.[8] Note that the boundary seemed to increase beyond the 35 dB burst attenuation, although the difference between this condition and the burstless condition fell short of significance in a t-test.

The discrimination (i.e., burst detection) results for the same stimuli are plotted in terms of percent correct as the filled circles in Figure 6. Performance was perfect for the original burst and declined with increasing

burst attenuation, first slowly, and then more rapidly beyond 25 dB. For stimuli with the initial [s]-noise, performance reached chance at the 35 dB attenuation. Note that the category boundary in the identification task continued to shift beyond that point for at least some listeners, suggesting that subjects' sensitivity to the burst was at least as great in phonetic labeling than in auditory discrimination. This result provides strong evidence of the sensitivity of phonetic categorization processes to very subtle changes in acoustic information.

Figure 6 also shows that burst detection was somewhat improved when the initial [s]-noise was removed, but only at the two weakest burst intensities (not a significant difference). Thus there may have been a slight auditory masking effect of the fricative noise on the burst, in agreement with Experiments 3 and 4.

Experiment 6

The purpose of Experiment 6 was to demonstrate a trading relation between burst amplitude and closure duration for the perception of a labial stop consonant. Labial bursts are weaker than alveolar and velar bursts (Zue, 1976), and informal observations have suggested that they are generally insufficient cues for stop manner. In other words, some closure silence is usually needed to perceive "sp," even with the original burst in place. This raises the question of whether labial bursts function as manner cues at all; perhaps, they merely add to the effective closure silence. Moreover, labial bursts offer the opportunity of observing not only effects of attenuation but also of amplification. Would an appropriately amplified labial burst become a sufficient stop manner cue?

The "slit"-"split" contrast was selected for the present study for several reasons. First, it has been used extensively in earlier studies (Bastian, Eimas, & Liberman, 1962; Dorman et al., 1979; Fitch et al., 1980; Marcus, 1978; Summerfield et al., 1981). Second, a "p" tends to be heard in this context as long as there are no strong cues to a nonlabial place of articulation in the signal portions surrounding the silent closure interval. That is, listeners report "split" when separately produced "s" and "lit" utterances are joined together with a sufficient interval of silence in between (Dorman et al., 1979). According to limited informal observations, the [l] resonances following a stop closure, unlike those of a full vowel, do not seem to harbor any significant formant transition cues to stop manner and place of articulation, which makes the "slit"-"split" contrast different from the "say"-"stay" contrast employed in Experiments 1-5. This fact may be partially responsible for the finding (cf. Fitch et al., 1980, and Best et al., 1981) that, in burstless stimuli, the typical "slit"-"split" boundary is located at much longer silent closure intervals (50-80 ms; for an exception, see Marcus, 1978) than the "say"-"stay" boundary (10-30 ms). Differences in place of stop articulation and in phonetic environment may also contribute to this boundary difference, however. One reason for conducting Experiment 6 (as well as Experiment 7) was to see whether the presence of a labial release burst, amplified to equal the power of an alveolar burst, might shift the "slit"-"split" boundary to the short silences characteristic of the "say"-"stay" boundary.

Method

Stimuli. A good token of "split" was selected from several utterances produced by a female speaker and was digitized at 20 kHz. In the original utterance, the initial [s]-noise (105 ms) was followed by a silent closure interval (about 150 ms) and a "blit" portion consisting of an initial release burst (16 ms), a voiced portion (about 230 ms), a silent [t]-closure, and a final [t]-release burst. The major energy of the labial release burst was concentrated in the first 4 ms. The rms amplitude of these first 4 ms was determined to be about 14 dB below the [l] maximum, and 20 dB below the [I] vowel maximum. The final 12 ms of the burst were about 13 dB below its initial 4 ms.[9]

Three additional stimulus versions were created either by amplifying or attenuating the 16-ms burst by 12 dB, or by eliminating it altogether. The (actual) silent closure duration in each of the four versions was varied from 40 to 100 ms in 10-ms steps. The resulting 28 stimuli were recorded in 10 different randomizations.

Subjects and procedure. The same ten subjects as in Experiment 3 identified the stimuli as "slit" or "split." Because the author noted that some of the stimuli sounded like "stlit" to him, this response alternative was provided as well. Stimuli without any clear consonant between the "s" and the "l" were to be considered instances of "slit."

Results and Discussion

Since "stlit" responses were rather infrequent, Figure 7 shows the combined percentage of "split" and "stlit" responses as a function of closure duration and of burst conditions. Three results are evident. First, attenuation of the burst by 12 dB had a clear effect, especially at the longer silences. Apparently, burst attenuation resulted not so much in a boundary shift as in a flattening of the labeling function. Second, the condition in which there was no burst at all gave results very similar to the attenuated-burst condition, provided the no-burst function is shifted to make the nominal closure durations comparable across the two conditions. (The actual closure durations were 16 ms longer, as indicated by the arrows in Figure 7.) This result is not surprising, given the initial low amplitude of the labial burst. Third, amplification of the burst by 12 dB had, surprisingly, no effect at all. One side effect of the amplification seemed to be a tendency to hear "stlit" rather than "split," in accord with recent data by Ohde and Stevens (1983) showing that burst amplitude is a cue to the labial-alveolar distinction. However, the present tendency was exhibited only by three of the ten subjects. A bias against the unfamiliar "stl" cluster may have played a role.

The effect of burst attenuation or elimination demonstrates that labial bursts, too, have a function as stop manner cues. The absence of any effect of burst amplification, however, suggests that the "slit"-"split" boundary cannot be easily pushed toward shorter values of silence. Although one might have expected burst amplification to shift the boundary on purely psychoacoustic grounds, it seems that the amplitude increment was either ignored by listeners or channelled into decisions about stop place of articulation rather than stop manner. This curious and potentially important finding called for a replication experiment.

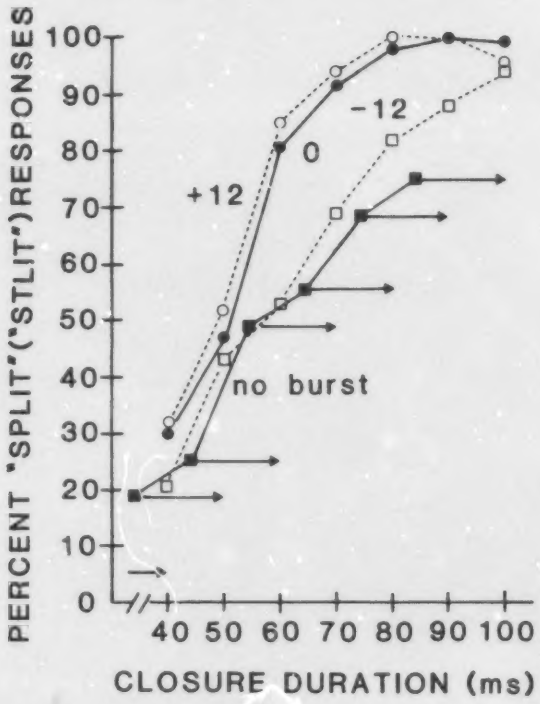


Figure 7. Effects of labial release burst amplitude in Experiment 6. Numbers refer to amplitude in dB relative to the original burst. Closure durations in the no-burst condition are nominal; actual durations are indicated by arrows.

Experiment 7

This study was similar to Experiment 6, except for differences in stimuli and the ranges of closure durations and burst amplitude values.

Method

Stimuli. Good tokens of the utterances "slash" and "splash" were recorded by a different female speaker and digitized at 20 kHz. The fricative noise of "slash" (142 ms) was used in all stimuli. The remainder (about 590 ms) was taken from "splash." This portion included an initial 10-ms release burst. (The original closure duration was 66 ms.) The amplitude of the burst was determined to be 7.4 dB below the [l] onset, 11.9 dB below the vowel maximum (75 ms later), and 2.9 dB above the fricative noise maximum (120 ms after noise onset). Six stimulus versions were created by leaving the burst unchanged, amplifying or attenuating it by 10 or 20 dB, or omitting it altogether. Each version occurred with (actual) closure durations ranging from 20 to 60 ms in 10-ms steps. The resulting 30 stimuli were recorded in 10 different randomizations.

Subjects and procedure. The same ten subjects as in Experiment 4 participated. They identified the stimuli as "slash" or "splash," with "stlash" as an additional option. To prepare the subjects for the amplified bursts, the instructions mentioned that some of the stimuli might have "pops" in them, which were to be ignored. The data of one subject had to be discarded because of numerous response omissions.

Results and Discussion

The results are shown in Figure 8. The left panel displays the labeling functions for the different burst amplitude conditions. In two respects, the findings replicate the principal results of Experiment 6: Attenuation of the burst necessitated a longer interval of silence, whereas burst amplification did not have the opposite effect; rather, amplified bursts seemed to function like slightly attenuated ones. In two other respects, the results are different from those of Experiment 6: The boundaries were considerably shorter here, and even the 20-dB burst attenuation condition still produced substantially more stop percepts than the burstless condition. These differences may indicate that the present release burst was a more powerful manner cue than that in the previous experiment. In addition, the different range of closure durations, as well as other stimulus characteristic, may have contributed to the boundary difference.

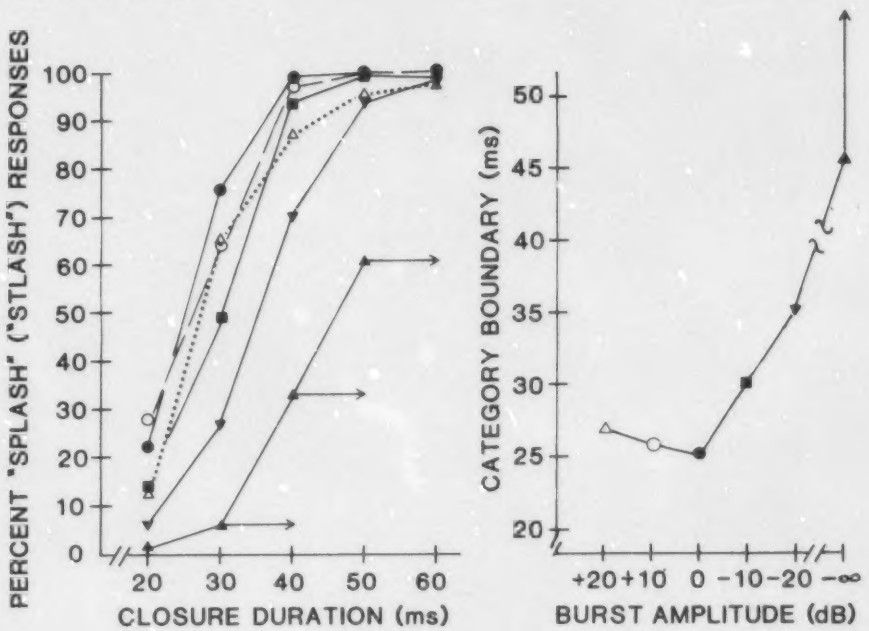


Figure 8. Effects of labial release burst amplitude in Experiment 7. Left panel is analogous to Figure 7, with legend provided by right panel. Right panel shows category boundary as a function of burst amplitude. Actual silence duration in no-burst condition is indicated by arrow.

The right-hand panel in Figure 8 summarizes the data by plotting the boundary location as a function of burst amplitude. It is plain that burst amplification did not continue the trend established by the burst attenuation results: As soon as the amplitude exceeded that of the original burst, its trading relation with silence duration came to an abrupt end.[10] How is this finding to be explained?

Only four of the nine subjects gave any "stlat" responses. These responses were fairly broadly distributed but tended to occur with the higher burst amplitudes and at short closure durations. However, these weak trends observed in a few subjects are not nearly sufficient to explain the sudden end of the trading relation between burst amplitude and silence duration.

A more relevant observation is that, to the author (and presumably to the subjects as well), the amplified bursts sounded like extraneous pops superimposed on the stimuli. This subjective impression suggests that amplification of the burst destroyed its auditory coherence with the other signal portions and caused it to "stream off." If so, it is particularly interesting that subjects perceived these stimuli not as if they had no bursts at all, but rather as if they had a burst of "normal" amplitude (see Figure 8). This finding thus seems related to two other intriguing phenomena described in the literature: duplex perception (e.g., Liberman, Isenberg, & Rakerd, 1981) and phoneme restoration (e.g., Samuel, 1981).

In duplex perception, a component of a speech stimulus is heard as a separate nonspeech event while, at the same time, it contributes to phonetic perception. Although the auditory segregation of the component is commonly achieved by dichotic channel separation, monaural duplex perception may occur when an acoustic cue, because of certain extreme properties, loses its coherence with the rest of the stimulus (see also Miller, Connine, Schermer, & Kluender, 1983). The present experiment seems to provide such an instance. Its results are also related to phoneme restoration, which is said to occur when a portion of a speech signal is replaced with an extraneous sound without affecting phonetic perception. Samuel (1981) has shown that, for restoration to occur, the extraneous sound must be a potential masker of the replaced portion. Thus, the so-called phoneme restoration effect may, at least in part, be a "cue restoration effect"; that is, listeners fill in missing acoustic information. A particularly relevant study was conducted by Pastore, Szczesiul, Rosenblum, and Schmuckler (1982): A syllable-initial [p] in one ear was perceived as "t" when a noise burst occurred in the other ear, but only when the noise included the frequencies typical of [t] release bursts. These findings combine aspects of duplex perception and cue restoration, as indeed do the present results. The amplified bursts were, of course, the best possible maskers of spectrally identical "normal" bursts, and because they segregated as "pops" from the rest of the signal, listeners were led to restore the original burst perceptually. If this interpretation is correct, then the data provide a particularly interesting demonstration of the detailed tacit knowledge of acoustic (or, perhaps, articulatory) properties of speech that listeners possess and apply in the course of phonetic perception.

General Discussion

The present series of studies fills some gaps in our knowledge of the acoustic cues for stop manner perception. They uniformly show that the release burst is a highly important cue for the perception of stops after [s].

One result that emerges from the experiments is that a natural alveolar release burst is usually sufficient to cue perception of a stop in the absence of closure silence (Exps. 1 and 5), whereas a natural labial release burst is usually not sufficient by itself (Exps. 6 and 7). Although, in the present studies, alveolar release bursts were followed by pronounced vocalic formant transitions while labial bursts were not, preliminary observations indicate

that the generalization holds regardless of following context, and that velar release bursts are similar in salience to alveolar ones. The greater power of alveolar and velar bursts is, in large part, due to their greater amplitude and longer duration, although spectral composition and/or different perceptual criteria for stops at different places of articulation may also play a role.

A second result of the present research is that listeners are extremely sensitive to the presence of even very brief or severely attenuated release bursts (Exps. 1, 2, 5). Experiment 5 showed that, when labeling stimuli phonetically, listeners are at least as sensitive to the presence of such minimal bursts as they are in a low-uncertainty burst detection task. As Nooteboom (1981) has pointed out, "phoneme identification seems to be an excellent way of measuring just noticeable differences" (p. 149). This is not a trivial result, for it suggests that the perceptual criteria employed in phonetic identification are extremely stable and finely tuned, despite the high stimulus uncertainty prevailing in a randomized identification test. Indeed, preliminary data suggest that this stability and sensitivity is maintained even in listening to fluent speech. The operation of stable criteria, internal to the listener and presumably shaped by language experience, is a hallmark of phonetic perception. Nevertheless, these criteria must also be flexible to accommodate natural variability in speech, such as might be due to changes in articulatory rate. In other words, the criteria are stable but not fixed; they are stable in the sense that their variability is not random but controlled by relevant factors.

A third finding is that release bursts, when shortened or attenuated in various degrees, engage in a regular trading relation with closure duration, a second important cue for stop manner: The weaker the burst, the more silence is needed to perceive a stop. There are two contrasting hypotheses about the origin of such a trading relation: It may either be phonetic or psychoacoustic in origin. According to the phonetic hypothesis (see Repp, 1982), the listeners' internal criteria specify the "prototypical" acoustic properties for the relevant phonetic segments, so that a reduction in one relevant property must be compensated for by an increase in another property to maintain the same response distribution. (A similar prediction could be derived from the information integration model of Oden and Massaro, 1978; see also Massaro and Oden, 1980.) According to the psychoacoustic hypothesis, on the other hand, the principal cue for stop manner resides in the onset characteristics of the signal portion (which includes the burst) following the closure silence, and the role of the silence is to prevent a forward masking effect of the preceding fricative noise on the auditory representation of those characteristics, and/or to enable the listener to attend to the critical onset properties. (This hypothesis is also congenial to the acoustic invariance hypothesis of Stevens and Blumstein, 1978.)

The present results are not wholly incompatible with psychoacoustic explanations. For example, the finding that attenuation of the fricative noise resulted in a reduction of the amount of silence needed for stop perception (Exps. 3 and 4), but only when a burst was present (Exp. 4), could be attributed to auditory forward masking. Effects of burst amplitude on stop manner perception also lend themselves to a psychoacoustic interpretation in terms of burst detectability. Data from other recent studies, however, argue strongly against a psychoacoustic account at least of the role of silence in stop manner perception. Best et al. (1981) found that the trading relation between closure duration and the F1 transition for the "say"-"stay" contrast

was absent in nonspeech analogs of the stimuli. Repp (1983b) demonstrated that this same trading relation, as well as that between closure duration and burst amplitude in "slit"-"split," was restricted to the phonetic boundary region but absent within phonetic categories. Perhaps the strongest result was recently reported by Pastore, Szczesiul, Rosenblum, and Schmuckler (1983): When the [s]-noise and the vocalic portion of "slit"-"split" tokens were differentially lateralized, so as to reduce peripheral auditory masking and facilitate selective attention, the amount of closure silence needed to perceive "split" remained the same. These results strongly favor a phonetic account of the integration of acoustic cues in stop manner perception, without ruling out certain psychoacoustic interactions in the peripheral auditory system that may, for example, affect burst detectability.

Two findings were unexpected and should provide a stimulus for further research. One result is that, apparently, burst amplitude has its effect on stop manner perception in absolute terms, not relative to the amplitude of the following signal portion (Exps. 3 and 4). The role that potential stop manner cues in this voiced portion may have played needs to be examined in a more controlled fashion. The results may suggest, however, that important stop manner cues reside in the first few milliseconds following the closure—that is, in the absolute magnitude and slope of the sudden energy increment.

A second unexpected finding was the absence of a trading relation between amplified labial release bursts and closure duration (Exps. 6 and 7). This phenomenon was tentatively interpreted as an instance of "cue restoration": The amplified burst was perceived as an extraneous "pop" and thus, instead of functioning as a cue in the speech signal, assumed the role of a masker for the cue expected by listeners—viz., of the "normal" release burst represented in listeners' detailed tacit knowledge of the normative acoustic properties of speech. A relation may exist between this phenomenon and the demonstration by Pols and Schouten (1978) that burstless initial stop consonants are more accurately perceived when preceded by pink noise (a potential masker of an absent burst) and Samuel's (1981) findings on the role of "bottom-up confirmation" in the phoneme restoration paradigm.

In conclusion, the present experiments have yielded factual information on the perception of a little-investigated cue as well as several intriguing effects that should stimulate further research. The results provide a modest challenge to psychoacoustic theories of speech perception. From a psychoacoustic viewpoint, stop manner perception seems a much simpler problem than, for example, perception of place of articulation: All that may be involved is the detection of some critical amount of energy increment or discontinuity in the signal. The eventual success or failure of psychoacoustic theories will rest, of course, on their ability to explain all kinds of phonetic perception, as well as to predict specific results from a model of auditory speech processing. Interesting work along these lines is now in progress (Delgutte, 1980, 1982; Goldhor, 1983), and the present data, being relatively straightforward, may provide a convenient testing ground for new models of peripheral auditory processing.

References

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. Journal of Experimental Psychology: Human Perception and Performance, 6, 536-563.

Bastian, J. (1959). Silent intervals as closure cues in the perception of stop phonemes. Speech Research and Instrumentation (Quarterly Progress Report, No. 33, Haskins Laboratories).

Bastian, J. (1962). Silent intervals as closure cues in the perception of stops. Speech Research and Instrumentation (Final Report [No. 9], Haskins Laboratories).

Bastian, J. Eimas, P. D., & Liberman, A. (1962). Identification and discrimination of a phonemic contrast induced by intervals of silence. Speech Research and Instrumentation (Final Report [No. 9], Haskins Laboratories).

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 29, 191-211.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. Journal of the Acoustical Society of America, 68, 843-857.

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In R. Carlson & B. Granström (Eds.), The representation of speech in the peripheral auditory System (pp. 131-149). Amsterdam: Elsevier.

Dorman, M. F., Raphael, L. J., & Isenberg, D. (1980). Acoustic cues for a fricative-affricate contrast in word-final position. Journal of Phonetics, 8, 397-405.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 65, 1518-1532.

Fant, G. (1973). Stops in CV-syllables. In Speech sounds and features (pp. 110-139). Cambridge, MA: MIT Press.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. Perception & Psychophysics, 27, 343-350.

Goldhor, R. (1983). The representation of speech signals in a model of the peripheral auditory system. Journal of the Acoustical Society of America, 73 (Suppl. No. 1), S4 (A).

Gurlekian, J. A. (1981). Recognition of the Spanish fricatives /s/ and /f/. Journal of the Acoustical Society of America, 70, 1624-1627.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics, 30, 133-143.

Marcus, S. M. (1978). Distinguishing 'slit' and 'split'—an invariant timimg cue in speech perception. Perception & Psychophysics, 23, 58-60.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. Journal of the Acoustical Society of America, 67, 996-1013.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. Journal of the Acoustical Society of America, 73, 2124-2133.

Nooteboom, S. G. (1981). Speech rate and segmental perception or the role of words in phoneme identification. In T. Myers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech (pp. 143-150). Amsterdam: North-Holland.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. Psychological Review, 85, 172-191.

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. Journal of the Acoustical Society of America, 74, 706-714.

Pastore, R. E., Szczesiul, R., Rosenblum, L. D., & Schmuckler, M. A. (1982). When is a [p] a [t], and when is it not. Journal of the Acoustical Society of America, 72 (Suppl. No. 1), S16 (A).

Pastore, R. E., Szczesiul, R., Rosenblum, L. D., & Schmuckler, M. A. (1983). Does silence play a peripheral role in the perception of stops? Journal of the Acoustical Society of America, 73 (Suppl. No. 1), S52 (A).

Pols, L. C. W., & Schouten, M. E. H. (1978). Identification of deleted consonants. Journal of the Acoustical Society of America, 64, 1333-1337.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H. (1983a). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. 10). New York: Academic.

Repp, B. H. (1983b). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Communication, 2, 341-361.

Repp, B. H., & Mann, V. A. (1980). Perceptual assessment of fricative-stop coarticulation. Haskins Laboratories Status Report on Speech Research, SR-63/64, 93-120.

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 7, 1124-1131.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

Summerfield, Q., Bailey, P. J., Seton, J., & Dorman, M. F. (1981). Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split.' Phonetica, 38, 181-192.

Zue, V. W. (1976). Acoustic characteristics of stop consonants: A controlled study. Lincoln Laboratory Technical Report, No. 523 (Lexington, MA).

Footnotes

[1]Dorman et al. (1980) found that the presence of an alveolar release burst was not sufficient for perception of a stop in vowel-fricative context (i.e., of an affricate, as in "ditch") in the absence of closure silence. While it is difficult to generalize from results obtained with single tokens of natural speech, it is possible that release bursts are more effective stop manner cues in fricative-vowel than in vowel-fricative environments.

[2]Amplitude measurements were performed after redigitizing the utterance without preemphasis, using a program of the ILS speech analysis system. The powerful appearance of the burst in Figure 1 is in part due to high-frequency preemphasis. "Vowel onset" refers here to the 20 ms of waveform immediately following the 20-ms burst. The burst, as defined here, may have included a first, extremely weak glottal pulse (between cutpoints 5 and 6 in Figure 1). No attempt was made to distinguish between transient, fricative, and aspirative phases of the burst (see Fant, 1973).

[3]Playback amplitude was not precisely calibrated but was held constant within a few dB by maintaining a certain setting of the level control on the tape recorder (Ampex AG500) for all subjects. The peak amplitude of the vowel at the subjects' earphones (approximately 83 dB SPL) was estimated postexperi-

mentally by converting the peak deflection of a vacuum tube voltmeter in response to the test syllables into dB SPL, according to a chart prepared by Haskins Laboratories technicians.

[4]For no particular reason, the burst was excised rather than infinitely attenuated. The latter procedure would have been preferable, but there are no serious consequences for the interpretation of the results. (The same applies to Experiments 6 and 7.)

[5]The reason for the different effects of vowel attenuation in Experiments 3 and 4 is not clear; they may have been due to different strengths of the stop manner cues in the vocalic portions used. In Experiment 3, no tendency to hear consonants other than "t" was noted, and a 40-ms silence always yielded close to 90 percent "stay" responses. In Experiment 4, on the other hand, a significant number of "svay" and "spay" responses occurred, and even when these were pooled with "stay" responses, the total percentage for burstless stimuli with a 40-ms silence was only 72. Therefore, the vocalic portion in Experiment 4 seemed to contain weaker stop manner cues than that in Experiment 3, and this may explain the different effects of attenuation.

[6]Finally, the pattern of "svay" and "spay" responses may be considered (Table 1). Attenuation of the burst increased both types of responses, simultaneously decreasing "stay" responses. Total elimination of the burst increased primarily "spay" responses. There was also a consistent Vowel effect, with both "svay" and "spay" responses being less frequent when the vocalic portion was attenuated. Fricative amplitude, on the other hand, had no effect on these responses at all. Closure duration did play a role (not shown in Table 1): "svay" responses decreased as closure duration increased in stimuli with bursts, but increased (high vowel amplitude) or remained constant (low vowel amplitude) in burstless stimuli; "spay" responses showed a strong increase with closure duration, provided they occurred at all (stimuli with low burst and high vowel amplitude, and burstless stimuli). The latter trend is in agreement with earlier observations that long closure durations favor perception of a labial place of articulation (Bailey & Summerfield, 1980). "Svay" percepts, on the other hand, may have resulted from either "misinterpreting" the burst as frication when the closure was short, or--in burstless stimuli--they may have taken the place of a possible "sthay" category, which is difficult to perceive but corresponds to the informal observation that burstless "day" portions often resemble "they." In either case, however, attenuation of the vocalic portion favored "stay" over "svay" and "spay," which indicates a role of the vocalic onset envelope in this distinction.

[7]There is a long-standing controversy, familiar from the literature on categorical perception (see Repp, 1983a, for a review), about whether speech perception experiments should be concerned with what listeners <u>can</u> do in an optimal situation or with what they do under normal circumstances. Auditory thresholds are often assessed in highly practiced listeners after many hours of training. No strong claim is being made here that these optimal thresholds coincide with the limit of burst effectiveness in phonetic identification, although they obviously define a lower bound. Rather, the hypothesis tested here concerns the burst detection threshold for unpracticed listeners in a brief discrimination test, on the assumption that this threshold is more likely to match the threshold of burst effectiveness in identification. In any case, the hypothesis is that listeners' sensitivity in phonetic identification

is no worse than in overt burst detection; if it is better, we would have evidence that the subconscious processes of phonetic identification maximally exploit auditory sensitivities.

[8]Some subjects gave many "svay" and/or "spay" responses; the former occurred most often at intermediate burst amplitudes and silences, the latter at low burst amplitudes and long silences. For the purpose of group boundary determination, these responses were grouped with "stay" responses.

[9]Inspection of the unpreemphasized waveform suggested that a first, very low-amplitude glottal pulse may have been included in the burst as defined here.

[10]It seems likely that amplification of the burst by just a few dB would still have increased its power as a manner cue. However, the present data suggest that the trading relation with silence duration ends well before a 10-dB gain is reached.

A PERCEPTUAL ANALOG OF CHANGE IN PROGRESS IN WELSH*

Suzanne Boyce+

Standard Literary Welsh exhibits a phenomenon known as "initial mutation," in which a lexical item may retain the initial consonant of its citation form or undergo one of three different rules that change its initial consonant by one feature. These rules are traditionally known as the SOFT, NASAL, and ASPIRATE mutations. The SOFT mutation changes voiceless stops and liquids to their voiced counterparts, and changes voiced stops to homorganic voiced fricatives. The ASPIRATE mutation changes voiceless stops to homorganic voiceless fricatives. The NASAL mutation changes voiced and voiceless stops to nasals but maintains voicing and aspiration characteristics (Fynes-Clinton, 1913). Examples (1)-(4) below illustrate, in order, the words /pot/ 'pot' and /beik/ 'bicycle' in CITATION form, and SOFT, ASPIRATE, and NASAL mutations.

(1a) (CITATION)	[ə pot]	(1b) [ə beik]
	The pot.	The bicycle.
(2a) (SOFT)	[ei bot]	(2b) [ei veik]
	His pot.	His bicycle.
(3a) (ASPIRATE)	[ei fot]	(3b) [ei beik]
	Her pot.	Her bicycle.
(4a) (NASAL)	[və m̥hot]	(4b) [və meik]
	My pot.	My bicycle.

The mutations are triggered by a preceding word or a particular syntactic context rather than phonological environment. Triggering contexts are idiosyncratic and dissimilar; typical contexts for the SOFT mutation, for

---

+Also Yale University.

Acknowledgment. This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am particularly grateful to Frank Gooding of the Linguistics Department at the University College of North Wales, Bangor, for sharing his time and laboratory, and to Carolyn Iorwerth for her advice and aid in finding subjects. Many thanks go to the Welsh-speaking students of University College, Bangor, and to the members of the Welsh club at Cambridge University for their voluntary participation in a somewhat frustrating experiment. I wish to thank my collaborators Cathe Browman and Louis Goldstein, and my colleagues Rena Krakow, Sharon Manuel, Harriet Nagen, Doug Whalen, Andre Cooper, Karen Kay, and Margaret Dunn for reading and commenting on innumerable drafts of this paper. Special thanks are also due to Gwen Awbery, for her comments on an earlier version.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

instance, are after the word i 'to,' adjectives after feminine nouns, and negative verbs that begin with [b],[d],[g]. Strictly speaking, therefore, conditioning for the mutations is neither morphological, syntactic, nor phonological, but something of all three.

There is a certain amount of converging evidence that the ASPIRATE and NASAL mutations are used less and less frequently in the spoken language. Jones (1977), for instance, states that "the Aspirate Mutation after other words [than ei 'her'] is rarely heard in spoken Welsh" (p. 105) and that "there is a tendency in many areas to use the Soft Mutation rather than the Nasal after ən ['in'] (p. 331). The most detailed analysis of this trend may belong to Awbery (in press), who presents evidence from a number of Southern dialects that the SOFT mutation is gaining ground at the expense of the ASPIRATE and NASAL mutations. Thus, in environments where Standard Welsh would use the NASAL or ASPIRATE mutation, Southern dialects may substitute the SOFT mutation. In addition, she notes a number of environments where the ASPIRATE mutation is dropped in favor of the CITATION form. Awbery states that dialects may differ as to which environments and which lexical items undergo the change, and that these changes are more common for younger than older speakers. Her examples for the (Standard) citation forms /ka:nol/, /kləwed/, and /karɛg/ are given below. The mutation being applied is in parenthesis.

(5a) (NASAL-Standard form)	[əŋ ŋha:nol ə ɬaur]
(5b) (SOFT-Dialectal form)	[əŋ ge:nol ə ɬaur]
	In the middle of the floor.
(6a) (ASPIRATE-Standard form)	[(ni) χləwes i ðɪm]
(6b) (SOFT-Dialectal form)	[gləwes i ðɪm]
	I didn't hear.
(7a) (ASPIRATE-Standard form)	[buru a χarɛg]
(7b) (CITATION-Dialectal form)	[buru a karɛg]
	To hit with a stone.

Awbery's claim is that although these changes show considerable variation among dialects and speakers, there is a clear pattern of change in progress from a four-way to a two-way system.

Given that such a change is occurring, the everyday experience of mutation for speakers in the South must be somewhat varied; that is, speakers must be accustomed to hearing both Standard and dialectal forms in the relevant mutation contexts. From the standpoint of any one speaker's experience, and regardless of whether the speaker's own grammar and productions are based on Standard or dialectal forms, the recognition system must anticipate alternative possibilities for those contexts of NASAL and ASPIRATE mutations in which substitutions may occur. In addition, overall, speakers must hear fewer instances of the NASAL and ASPIRATE mutations than of the SOFT mutation and CITATION forms. Presumably speakers are aware of this situation at some level of their internal grammar; that is, they must 'know' that the NASAL and ASPIRATE mutation contexts are problematic.

It is often hypothesized that language change coalesces around some point of vulnerability in the system (opacity, hole in the pattern, etc.). In this vein, it's interesting to note that even in Standard Welsh, there are many

more contexts that require the SOFT mutation than the ASPIRATE and NASAL mutations; an informal count of triggering contexts listed in Jones (1977) reveals 2 for NASAL and 10 for ASPIRATE, as opposed to 51 for SOFT. This imbalance has apparently (Warren Cowgill, personal communication) always existed in the history of Welsh; although all three mutations have been steadily losing contexts, the NASAL and ASPIRATE mutations have always been relatively the most impoverished. (Note, however, that both NASAL and ASPIRATE mutations occur in some very common phrases—for instance, the NASAL mutation after ən 'in' and the ASPIRATE mutation after ei 'her.' This may mean that if one factor affecting vulnerability to linguistic change is frequency of usage, the particular measure applied must be based on regularity of usage, or number of forms subject to the rule, rather than simple text frequency.) Thus, historical data as well as data from current productions in spoken Welsh suggest that the ASPIRATE and NASAL mutations are weakening.

The results we present in this paper are focused on the state of the mutation system as a result of the change in progress in spoken Welsh. However, these data are derived from a series of experiments originally designed to speak to a different issue, that of the internal structure of the lexicon for morphologically related words. Because of this separation between the original aim of the experiment and the way we look at the data here, only a very brief description of the experimental design and the theory behind it is offered below. The entire series of experiments is reported in detail in Boyce, Browman, and Goldstein (in preparation).

Briefly, the experiments involved a method known as repetition priming. This technique relies on the fact that a subject who has heard or read a word recently will recognize it faster and more accurately when it is presented a second time, that is, the subject is "primed" for recognition of that word. The effect has recently been manipulated to probe the organization of the lexicon for morphologically related words by testing which pairs of related words produce a priming effect (Stanners, Neiser, Hernon, & Hall, 1979). Thus, our experiments were structured to measure priming between various forms (CITATION, SOFT mutation, etc.) of the same lexical item.

Procedure

Subjects listened first to a list of "priming" words and then to a second list of "target" words that were obscured by simultaneous random noise.[2] Words used were mono- and bi-syllabic masculine nouns beginning with a voiced or voiceless oral stop, and were carefully balanced for number of (citation form) initial /p/,/t/,/k/,/b/,/d/,/g/. All had stress on the first syllable.

Each word was presented in a syntactic context that required a particular mutation and was recorded onto tape by a native speaker of North Welsh. The contexts were as illustrated in examples (1)-(3) above with the addition of optional postpositions: (a) ei __ o (SOFT MUTATION); (b) ei __ hi (ASPIRATE MUTATION); and (c) ə __ əma (CITATION FORM for masculine nouns). The postpositions mean, in order, 'of him,' 'of her,' and 'here' or 'this.' (All three phrases are in current colloquial usage.) Subjects were told which phrases would occur and were asked to write each phrase in full if they could. Only full phrases with correct context and mutated form as well as correct lexical item were scored as correct responses. Note that although in general the ASPIRATE mutation is subject to dialectal substitution, the ASPIRATE mutation in the context "ei __ (hi)" is rigorously observed (Jones, 1977, p. 105).

Presumably this is due to contrast with the SOFT mutation context "ei __ (o)." To simplify experimental design, the NASAL mutation was not used.

Subjects

The subjects were 60 native speakers of Welsh recruited through Welsh-speaking clubs at the University of Bangor, Wales, and Cambridge University, England. Of these, 34 were born and educated in North Wales, and 26 were born in South Wales. (The major dialect boundary for Welsh runs between South and North Wales.) Forty-eight of the subjects had experience with Northern dialect from living in North Wales; the other 12 (all born in the South) were accustomed to Northern dialect from radio programs and friends. None had any difficulty understanding the Northern pronunciation of the speaker who made the tape.

Results

We present data here from two experiments. As noted above, both experiments were set up to contrast different prime-target combinations for the three contexts used; however, for our purposes here the relevant comparison is always, for instance, all CITATION form means versus all ASPIRATE mutation means, all ASPIRATE mutation means versus all SOFT mutation means, and so on.

In Experiment 1, each form of a lexical item (CITATION, SOFT mutation, or ASPIRATE mutation) was primed by that lexical item in the same form. This was contrasted with conditions in which the target word was not primed. In all, eight different lexical items (words) were used. Each was represented once in the appropriate form (CITATION, SOFT mutation, ASPIRATE mutation) in each condition. The following table shows the results as mean percent correct responses to the target form.

	PRIME:	SELF	NONE
T A R G E T	CITATION FORM	73	46
	SOFT MUTATION	67	46
	ASPIRATE MUTATION	36	20

Here we see that means for the CITATION form and SOFT mutation are nearly identical in both the SELF and NONE conditions. This means that the CITATION and SOFT mutation forms behave similarly under both presentation conditions. In contrast, the means for the ASPIRATE mutation are considerably lower. (Analysis of variance showed the difference between the three sets of means to be significant at the 1% level. A posteriori contrasts between pairs of means for the CITATION, SOFT mutation, and ASPIRATE mutation indicate this difference is due to the lower ASPIRATE mutation means. There was no significant difference between means for the CITATION form and SOFT mutation.) This suggests that, even when subjects had been previously exposed to the same word, in the same mutating phrase, words to which the ASPIRATE mutation had applied were more often misperceived. This difference between SOFT and ASPIRATE muta-

tions held for speakers born in both the South and North regions. (The interaction between mutations and area of speaker was not significant.) Rescoring in which all phrases with the correct lexical item were counted (regardless of mistakes in context or mutation heard) did not alter these results. Thus, the weakness of the ASPIRATE mutations does not represent some "bias" on the part of subjects against reporting the ASPIRATE mutation, or against reporting the "ei __ hi" 'hers' context. Rather, actual recognition of the lexical item is impeded in this context.

The second experiment is essentially a replication of the first, for a larger word set and with the addition of two more prime-target conditions. The CITATION form was excluded. Thus, forms in both mutations were presented in the following conditions: (i) primed by themselves (SELF priming); (ii) not primed; (iii) primed by the citation form (BASE priming); and (iv) primed by the other mutation (OTHER priming). The following table shows the data for the SOFT and ASPIRATE mutation under each of these conditions, again as mean percent correct recognition of the target mutated form. This time, 32 lexical items were used. Again, each appeared once in each condition.

	PRIME:	SELF	NONE	BASE	OTHER
TARGET	SOFT MUTATION	61	41	49	49
	ASPIRATE MUTATION	47	27	43	39

As in Experiment 1, in all conditions those forms to which the ASPIRATE mutation has applied are poorly recognized compared to forms in which the SOFT mutation has applied. (Analysis of variance showed the difference between the two sets of means to be significant at the 3% level.) Again, this pattern holds for speakers from both regions (the interaction of area by mutation was not significant), and rescoring again made no difference.

Discussion

Taken together, these results parallel the change in the mutation system documented by Awbery. Her evidence shows the weakness of the ASPIRATE mutation, as a rule that is being replaced by another rule, and suggests that the CITATION form and the SOFT mutation contrast with the ASPIRATE as lively, well-established rules in the grammar of Welsh. The experiments described above show that this linguistic situation is reflected in (1) an equal probability that the CITATION and SOFT mutation forms will be correctly identified and (2) a greater likelihood that forms in the ASPIRATE mutation will be misperceived or missed.[3] This result is particularly striking because, as noted above, the context "ei __ hi" is an extremely robust environment for the ASPIRATE mutation. Thus, the differential effect for the ASPIRATE mutation occurs in a context exempt from the change in progress. This shows that it is the rule <u>itself</u>, with all the contexts in which it applies, that is problematic rather than one particular syntactic or morphological context. Further, speakers from both dialect areas show this effect of decreased perceptibility for the ASPIRATE mutation.

These results are interesting for several reasons. First, of course, our experiment constitutes independent and empirical support for Awbery's hypothesis about mutation rule change in Welsh. More importantly, our experiment shows that rule change in progress may be reflected in a tendency to confuse or misperceive input that is eligible to undergo the changing rule. We have seen that for all speakers, regardless of dialect area, any ASPIRATE mutation context is susceptible to misperception. This is clear because the misperception occurs even in the robust context 'ei __ hi" 'hers,' which is not subject to dialectal substitution or change. It is not clear how much this decrease in perceptibility for the ASPIRATE mutation is due to the subjects' experience of dialectal substitution in ASPIRATE mutation contexts, and how much to internal, grammar-related factors that may have led to changing productions in the first place. We know that (many) Southern speakers are accustomed to experiencing an unstable situation for the ASPIRATE mutation, but data on how much the experience of Northern speakers includes substitutions in ASPIRATE mutation contexts are currently unavailable. Production data from the North parallel to Awbery's are needed to sort out these possibilites. It is possible that a study of Northern dialects would reveal a similar pattern of change in progress. If so, then the interpretation of our data is the same for both Northern and Southern speakers, i.e., that the recognition system changes as production changes---in some cases, as in our robust 'hers' context, it may even anticipate production for environments that are eligible to undergo the changing rule, but don't. On the other hand, if no such changes are reported in Northern dialects, our experiments may have tapped the early stages of a change that has not yet emerged into production in the North.

References

Awbery, G. M. (in press). Moves towards a simpler, binary mutation system in Welsh. In H. Andersen & J. Gvozdanovic (Eds.), Workshop on Sandhi phenomena in the languages of Europe (International Congress of Phonetic Sciences, Utrecht, The Netherlands, August 1-6, 1983).

Boyce, S. E., Browman, C. P., & Goldstein, L. M. (1984). The organization of the lexicon for Welsh: An auditory priming study. Unpublished manuscript.

Fynes-Clinton, O. H. (1913). The Welsh vocabulary of the Bangor District. Oxford: Oxford University Press.

Jones, T. J. R. (1977). Living Welsh. Suffolk: Hodder and Stoughton.

Kempley, S. T., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. British Journal of Psychology, 73,441-454.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. Journal of Verbal Learning and Verbal Behavior, 18, 399-412.

Footnotes

[1]More precisely, it voices [p],[t],[k],[ʧ] and [r_o^h], and spirantizes [b],[d] and [m]. The fricative reflex of [g] was once realized as [ɣ] but has since disappeared.

[2]It has been shown (Kempley & Morton, 1982) that target words that have been primed are more readily recognized in noise.

[3]An alternative explanation for these data based on differences in discriminability for fricatives and stops may occur to the reader. Notice, however, that equal numbers of both phonetic categories appear in both mutations (e.g., [bot] vs. [veik], [fot] vs. [beik]), and that while words with initial voiced fricatives were somewhat better recognized than words with initial voiceless fricatives, words with initial voiceless stops were better recognized than words with initial voiced stops; thus, the effects should even out. In addition, a parallel experiment (not reported here) using words whose initial consonants are never subject to mutation showed the same differential effect in ASPIRATE mutation contexts. This evidence is examined in greater detail in Boyce et al. (forthcoming).

SINGLE FORMANT CONTRAST IN VOWEL IDENTIFICATION*

Robert G. Crowder+ and Bruno H. Repp

Abstract. Subjects rated ambiguous steady-state vowels from a continuum with respect to the categories /i/ and /ɪ/ (Experiment 1) or /ɛ/ and /æ/ (Experiment 2). Each target was preceded, .35 sec earlier, by one of the following precursors: (1) one endpoint from the target continuum, (2) the other endpoint, (3) the isolated first formant (F1) from (1), (4) the isolated F1 from (2), or (5) a hissing noise. Although (3) and (4) did not sound as if they came from the target continuum, they produced reliable contrast in both experiments. In the /i-ɪ/ experiment, single-formant contrast was as powerful as from the full vowels. These results suggest a sensory, rather than judgmental, basis for the vowel contrast effects obtained.

The occurrence of contrast in perceptual judgments along single dimensions is so commonlyplace it seems almost uninteresting. In judging shades of grey, heaviness of lifted objects, line lengths, loudness of tones, and so on, the perceived magnitude of one stimulus is usually affected contrastively by another stimulus with which it is presented. A patch of grey seems dark against a white background and yet the same patch seems light against a black background, for example. The pervasiveness of contrastive interactions between nearby stimuli should not, however, lead us to forget how important it is. Contrast allows the perceptual system to focus on what otherwise might be elusive differences. It hardly requires discussion that edge sharpening, in vision, advances the more informative aspects of the visual world at the expense of the less informative aspects.

The example of visual brightness contrast is an interesting one because a detailed neurophysiological basis for it has been worked out (however, see Gilchrist, 1977). Edge sharpening in at least simultaneous brightness contrast follows inescapably from verified rules of recurrent lateral inhibition in the visual (retinal) system (see summary in Lindsay & Norman, 1977). Where does this leave us with other kinds of contrast, though? It would be grandiose to apply the neural circuitry proposed for brightness contrast to, say,

---

*Also Perception & Psychophysics, in press.

+Also Yale University.

Acknowledgment. Part of the preparation of this manuscript was done while R. Crowder was a fellow at the Center for Advanced Study in the Behavioral Sciences. We are grateful to the following sources of support: NSF Grant No. BNS8206304 to the Center for Advanced Study in the Behavioral Sciences, NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories, and NSF Grant BNS 80005838 to R. Crowder. We appreciate the able assistance of Virginia Walters and Herta Flor in collecting and analyzing the data.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

contrast effects in judging the conservatism of Supreme Court Justices. Some kinds of contrast, in other words, might be more "cognitive" or judgmental and others more sensory. (This comment need not seem like an eruption of the mind/body distinction: Infants would display the sensory forms of contrast but not the judgmental forms.)

In fluent speech perception, most especially in unstressed, "reduced" vowels, the stimulus information is often impoverished relative to prototypical category instances. In a speaker's haste to get from one to another consonantal gesture, he/she often "misses" producing a vowel sound in anything like its citation form. Veridical perception would be well served, in these cases, by a process of "edge sharpening" for the vowel perception system, so that surviving acoustic stimulus distinctions would be exaggerated. Indeed, since the work of Fry, Abramson, Eimas, and Liberman (1962), it has been known that isolated vowels show contrastive context effects in identification judgments. Sawusch, Nusbaum, and Schwab (1980) have distinguished three classes of explanation for the various instances of vowel contrast that have been reported in the intervening years. They discount feature detector fatigue as an explanation because there are often substantial time lapses between context and target. We can also mention that retroactive contrast (Diehl, Elman, & McCusker, 1978; Repp, Healy, & Crowder, 1979)--where the target comes first and the context second--effectively dismisses this first explanation. Changes in auditory ground and response bias are the two remaining classes of hypothesis. Sawusch et al. (1980) apply these interpretations mainly to contrast elicited in an anchoring paradigm and we need not follow these applications here in detail. It suffices to remark that the auditory-ground interpretation appeals to sensory contrast in a way that is congenial with the analogy to visual brightness contrast, whereas response bias would very clearly be a judgmental process.

Some recent research on selective adaptation in the perception of stop consonants has pointed towards auditory-sensory explanations rather than towards judgmental explanations. In the selective adaptation paradigm, repeated presentations of an adaptor stimulus are shown to affect the perception of a subsequent test stimulus. The experimental operations involved in measuring selective adaptation are obviously a special case of contrast, several authors having suggested a profound continuity of process between the two (Crowder, 1981; Diehl et al., 1978; Diehl, Lang, & Parker, 1980). In two such experiments, the authors were able to pit sensory (spectral) and judgmental factors against each other. In one of these studies, Roberts and Summerfield (1981) used an audio-visual adaptor in which an acoustic /be/ was synchronized with a visual /ge/. The combination was identified as /de/ or /ðe/; however, its effect on perception of a /be-de/ test series was identical to that of an unambiguous acoustic /be/. Thus, perception responded to the spectral, and not the perceived phonetic, nature of the adaptor. In the other experiment (Sawusch & Jusczyk, 1981), an adaptor was made from a fricative-stop-vowel syllable /s+ba/ with 75 ms of silence between the two segments. Under these conditions, subjects call the adaptor syllable "spa" even though the stop-vowel portion, alone, is unambiguously /ba/. In a /ba-pa/ test series following these and other adaptors, the "perceptual /spa/" but "acoustic /s+ba/" affected responses just the same way as did an unambiguous /ba/, again showing the spectral cue to prevail even in the face of contradictory labeling.

In their experimental work with isolated vowels from the /i/-/ɪ/ continuum, Sawusch et al. (1980) reached the interesting conclusion that both sensory and judgmental factors contribute to contrast, but at different ends of the continuum. Their technique involved measurement of discrimination sensitivity as well as identification ratings. When /i/ was the context (anchor), there seemed to be genuine changes in the sharpness of the sensory system in the /i/ range of the continuum. When /ɪ/ was the context, however, the changes seemed rather to be in where people placed their response criterion. Thus, within the same stimulus continuum we may be able to see more than one form of contrast operating.

Experiment 1

The present report extends the Sawusch et al. work in several ways, but it explores the same question of where contrast effects should be located with regard to the sensory versus response end of the processing machinery. In our experiment, pairs of isolated vowel sounds were presented in rapid (.35 sec) succession. The stimuli all came from a seven-item /i/-/ɪ/ continuum varying only in F1 frequency. (The stimuli used by Sawusch et al. varied in F2 as well as F1.) The second item in each pair was the target (second through sixth item from the continuum) and the first was either of the two endpoint items, either of these two items with F2 and F3 removed, or a control (hiss). No response was required to the first item in each pair, the context item. These conditions were either mixed together randomly in a continuous series of trials or were presented in blocks. In the latter arrangement we should be in a position to observe the anchoring effects found by Sawusch et al. as well as "regular" contrast between the two items in a pair. When the conditions are randomized, however, only pairwise contrast should be observed. Our choice of the neutral hiss in the control condition was considered: We wanted as close to a "no contrast" condition as we could get. Any tone or vowel, however unrelated to the test continuum it seemed, carried potential spectral or phonetic bias. The hiss served as a simple warning signal with no such bias.

Because removal of F2 and F3 results in these items' sounding unlike tokens of the /i/-/ɪ/ continuum, we can also offer expectations for which contrast effects ought to be influenced by whether the precursors are intact three-formant vowels or not. If Sawusch et al. are correct in assigning contrast produced by /i/ to sensory factors, we might expect that removal of F2 and F3 would make little difference. For example, Crowder (1981, 1982) has proposed a theory of frequency-specific recurrent lateral inhibition (see below) that anticipates the same degree of contrast whether or not F2 and F3 are present. If contrast from the /ɪ/ side of the continuum is produced by other factors, perhaps response bias or a range-frequency effect (Parducci, 1974), then removal of F2 and F3 might alter the situation, because the tacit labels that subjects might assign to the precursors (and that might engage the judgmental bias) would be foreign to the target /i/-/ɪ/ continuum. Furthermore, if a sort of adaptation-level mechanism contributed to contrast in the anchoring situation (Sawusch et al., 1980), then we should expect more contrast in the blocked arrangement of conditions than in the randomized arrangement; this is because in the blocked arrangement the same single precursor is the first item in each pair and therefore vastly outnumbers each of the six items that can be the second item in the pair.

Method

Subjects. The subjects were 20 Yale undergraduate, serving either for pay or for course credit.

Stimuli. The basic continuum of seven vowels was prepared on the Haskins Laboratories Parallel Resonance Synthesizer. The items were designed to range perceptually from /i/ to /ɪ/ and varied only in F1 center frequencies (from 279 to 381 Hz in roughly equal steps). F2 and F3 center frequencies were kept fixed at 2075 and 2780 Hz, respectively. These frequencies were compromise values between those typical of the vowels /i/ and /ɪ/ (Peterson & Barney, 1952). Three additional stimuli were used: (1) the /i/ endpoint from the continuum with F2 and F3 removed (through options within the synthesizer); (2) the /ɪ/ endpoint modified in the same way, and (3) a soft hiss, which served as a control. All stimuli were 300 ms long. The vowels rose in fundamental frequency from 80 to 100 Hz during the first 100 ms and declined to 85 Hz during the last 100 ms. The amplitude envelope was likewise shaped at the beginning and end of the syllable. In the stimuli with F2 and F3 removed, the amplitude of F1 matched F1 amplitude in the corresponding full vowels. However, the overall amplitude was reduced by removal of F2 and F3.

In a preliminary experiment, 25 subjects were given single-item identification tests on vowels similar to the vowels lacking F2 and F3 used in the present experiment. Other details of that preliminary experiment need not concern us: It incluuded contrast comparisons similar to, but superceded by, those of the present experiment. Nothing in the preliminary study compromises, however, what we found later. Of interest now is that these 25 subjects were asked to listen to tokens of the various precursors in isolation and report what they sounded like, with examples of words containing the sounds. It has long been known that people can perceive and classify single-formant vowels, and that low frequency single formants are heard as back vowels (Delattre, Liberman, Cooper, & Gerstman, 1952). When describing the /i/ endpoint of the vowels described above, but with F2 and F3 deleted, 24 of the 25 subjects reported it sounded like /u/ (BOOT) and the remaining subject reported /ə/. Given the /ɪ/ endpoint, 7 responded with the same vowel (/u/), 14 with the sound /o/ (BOAT), 3 with /ʌ/, and 1 with /ə/, but never /ɪ/. Thus, we may be assured that removal of the second and third formants did indeed drive labeling away from the /i/-/ɪ/ continuum.

The experimental tapes contained 100 trials (pairs of vowels) apiece. Each trial included the precursor, a 350 ms delay, and then the target; after the offset of the target, there was a 2.5 sec delay before the beginning of the next trial. After every 10 trials, there was a longer intertrial delay (5 sec) intended to help subjects keep their places.

Procedure. Subjects in the blocked condition received five 100-trial tapes, one for each of the five precursor conditions, in different orders as determined by a Latin square. The 100 trials with a given precursor included 20 with each of the target vowels (numbers two through six on the original continuum). Subjects in the random condition heard precisely the same 500 trials, also in batches of 100; however, the trials were completely randomized, so that two adjacent trials usually had different precursors.

In the first part of the session, subjects were played the /i/-/ɪ/ continuum three times, in order, from number two through number six of the original seven. They were told that the first item in the group is "what we are calling EE" and the last is "IH." They all then listened to the first 10 tri-

als on the Random tape as practice, and then began the experiment proper. Answer sheets had arabic numbers from 1 to 5 opposite each trial number. Over the left column (the 1's) the word BEET was spelled and over the right column (the 5's) BIT. The procedure was to rate the similarity of each target to the vowels in these two prototypes by circling one of the five numerals.

Results

The main results are shown in Figure 1, in terms of mean rating in the "IH" direction. The left panel shows ratings when the precursors were the full-vowel endpoints from the continuum; the right panel shows what happened when F2 and F3 were removed from these stimuli. The Hiss condition is drawn in both panels as a baseline.

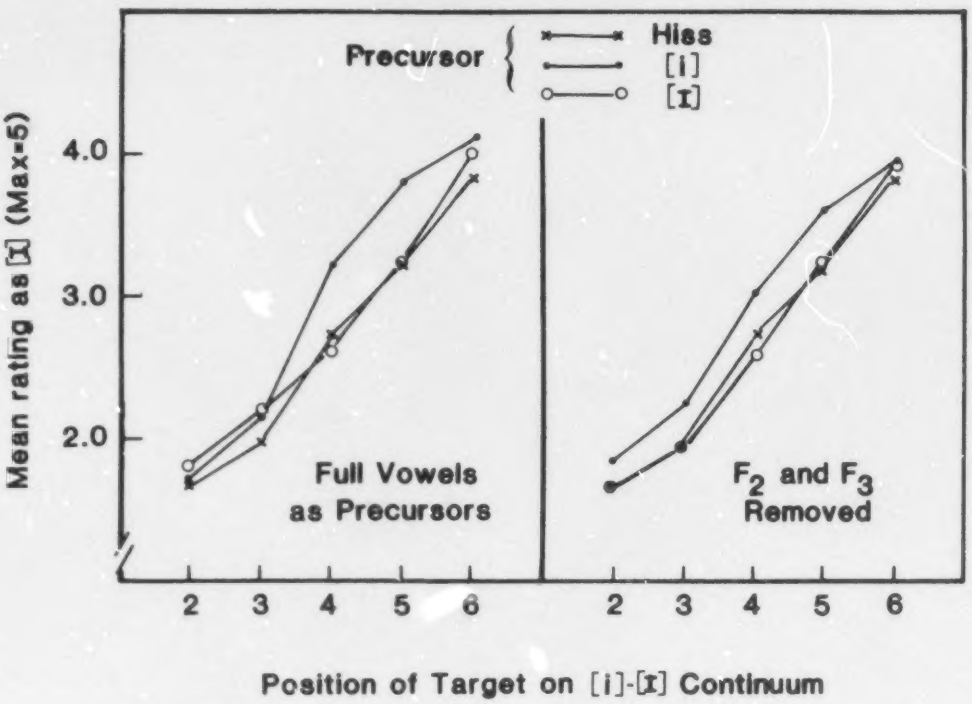


Figure 1. The effect of five context precursors on the relationship between the position of a target vowel along the /i-ɪ/ continuum and the tendency to rate it as /ɪ/. The same control condition with a hiss as context is plotted in both panels. On the left, data are shown when the precursor was one or the other of the two series endpoints /i or ɪ/. On the right are the results when the same precursors were used with all but the first formant deleted. (Experiment 1)

First of all, there was no result of blocking versus randomizing the experimental conditions and so Figure 1 combines these two conditions. In a 2 (blocked/random) X 5 (precursor) X 5 (target position on the continuum) analysis of variance, blocking did not approach the .10 alpha level, either as a main effect or in interactions. However, the same analysis of variance showed there were reliable differences among the five precursor conditions, $F(4,72) = 7.11$, $p < .01$, differences that interacted with target position, $F(16,288) = 3.41$, $p < .01$. As the figure shows, precursors had little or no effect on the relatively unambiguous targets--numbers 2 and 6 from the contin-

uum. Of course, the position on the continuum of the target itself had a reliable main effect on ratings of "IH-ness," $F(4,72) = 120.06$, $p < .01$.

Contrast was asymmetrical from teh two ends of the vowel continuum, to say the least: There was none when /ɪ/ or its single-formant version were used as precursors. However, the /i/ precursor quite obviously made the targets sound more like /ɪ/, and this was true with or without F2 and F3. The next question is whether the overall degree of contrast was different for the full vowels (left panel) versus the vowels with F2 and F3 removed (right panel). The best single measure of the contrast effect is probably the difference between the /i/ and /ɪ/ precursors (or their modified versions). In a new analysis of variance, the control condition was dropped, and the factors were (1) blocked/random, (2) full/altered precursors, (3) /i/ versus /ɪ/ precursors, and (4) position of the target on the continuum. There was a reliable main effect of whether the precursors were altered or not, $F(1,18) = 5.25$, $p < .05$, reflecting the fact that the full vowel precursors (left panel of Figure 1) led to somewhat higher /ɪ/ ratings whatever the identity of the precursor (/i/ or /ɪ/). However, the identity of the precursor had a large main effect, $F(1,18) = 13.91$, $p < .01$, and it did not interact with whether precursors were full or altered, $F < 1$. Thus, there was no evidence that full vowels exert a greater contrast effect than vowels with F2 and F3 removed. There was a statistically significant interaction between these two factors and the position of the target along the continuum, $F(4,72) = 3.23$, $p < .05$. This interaction reflects the fact that the full vowels had their effects exclusively on the fourth and fifth continuum positions, whereas the contrast produced by altered vowels occurred at all continuum positions except the last.

Discussion

The main finding of the experiment is that contrast was obtained even after F2 and F3 were removed from the endpoint vowels, rendering them phonetically foreign to the continuum being judged. The degree of contrast was not even changed by this operation. It is true that the contrast effect "bulged" differently with the full vowels than with the altered vowels, as revealed by the significant three-way interaction identified in the previous paragraph. We defer comment on this finding until after reporting the second experiment.

Another finding was the asymmetry in contrast across the /i/-/ɪ/ continuum. Whereas Sawusch et al. (1980) observed anchor effects from both ends of this continuum, and later were able to assign them to different mechanisms, we simply got no contrast at all from /ɪ/. At the very least, the asymmetry of this continuum in contrast tells us there is more at work here than a simple judgmental bias leading people to assign contrasting labels to precursor and target. Such a bias would result in symmetrical effects. In fact, if we accept the Sawusch et al. analysis, our results suggest that judgmental bias (associated with the /ɪ/ context) simply did not occur in our experiment.

It made no difference whether conditions were blocked or mixed randomly across trials. This is comforting in that it means there is one less choice to worry about in designing experiments. It was disappointing in the context of this experiment, however, for if contrast from /ɪ/ precursors were a consequence of judgmental bias, one might have expected blocking to make a difference, and differentially for the full and the altered vowels. One possibility is that judgmental bias was not engaged in the pairwise, precursor-target trial arrangment because the precursor never required an overt labeling response. In the anchoring literature, all items are identified in sequence. Perhaps

requiring people to label both of the two items in each pair (as in Repp et al., 1979) would have been sufficient to make blocking a more interesting variable. In addition, the short interval between the stimuli in a pair (350 ms) may have discouraged subjects from assigning covert labels to the precursors.

Single-formant contrast is predicted by Crowder's (1978, 1981, 1982, 1983) theory. The hypothesis is that vowels are represented in auditory memory in some form similar to a sound spectrogram. When two tokens are together in auditory memory, they show frequency-specific lateral inhibition; that is, where the two share formant energy, they mutually weaken each other's representation. This process is shown schematically in Figure 2 for an /a/-/æ/ pair.

The inhibition has no effect on vowel quality when formants match from the two vowels. However, when formants partially overlap, as in the illustration of Figure 2 or in F1 of nearby members of the /i/-/ɪ/ continuum, the intersection region will be inhibited in both. This leaves the most extreme regions of the intersecting formants intact, giving them more extreme formant center frequencies after inhibition than they had before. The absence of contrast from /ɪ/ tokens is, of course, as baffling to this model as it is to most others. At this point, we decided to replicate our experiment with another vowel continuum, in order to see whether these findings had any generality.

Experiment 2

The second experiment used vowels from the /ɛ/-/æ/ continuum and dropped the comparison of randomized and blocked conditions, using blocked presentation only. Otherwise, it was nearly identical to Experiment 1. The purpose was simply to generalize to different subjects and different vowels the occurrence of single-formant contrast.

Method

Subjects. The subjects were 30 Yale undergraduates, participating in return for pay.

Stimuli. Another seven-vowel continuum was prepared on the Haskins Laboratories Software Serial Synthesizer. (No parallel synthesizer was available to us at the time.) The vowels were designed to range perceptually from /ɛ/ to /æ/ and varied, this time, both in F1 and F2 (respectively, from 530 to 660 Hz and from 1840 to 1720 Hz); F3 was fixed at 2480 Hz. The response alternatives on the extremes of the fine-point rating scale were the words BET and BAT. Low-pass filtering (cutoff frequency = 800 Hz, rolloff = 48dB/octave) was used in order to produce endpoint tokens of /ɛ/ and /æ/ with F2 and F3 deleted. These altered versions of /ɛ/ and /æ/ sounded to us unambiguously like /ʌ/ and /ɔ/, respectively. The same hiss was used as in Experiment 1. The items were all 260 ms long, the ISI was .35 sec, and the delay between trials was set at 3.5 sec. In all other procedural details, this experiment was identical to the blocked condition of the previous one.

Results

The results are shown in Figure 3, which is organized identically to Figure 1. Evident in the figure are several findings: (1) Contrast from the /ɛ/ direction on the vowel continuum occurred both for the full vowel precursors

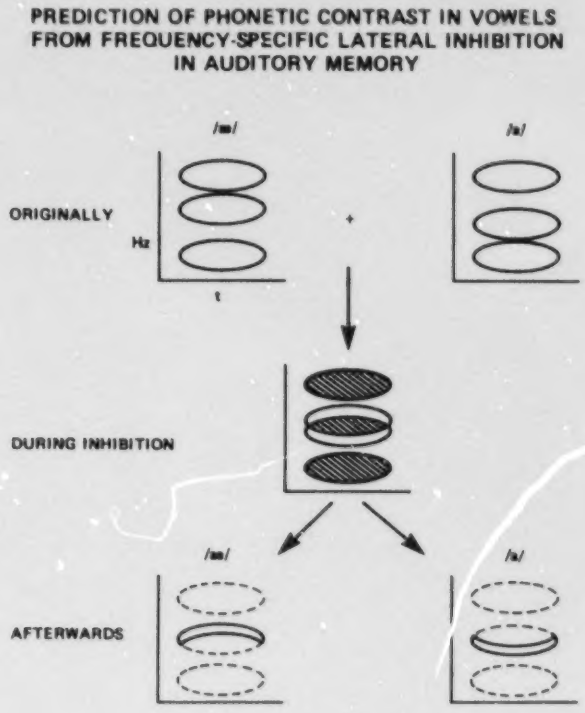


Figure 2. An illustration of how frequency-specific lateral inhibition could produce phonetic contrast in vowels. See text for explanation.

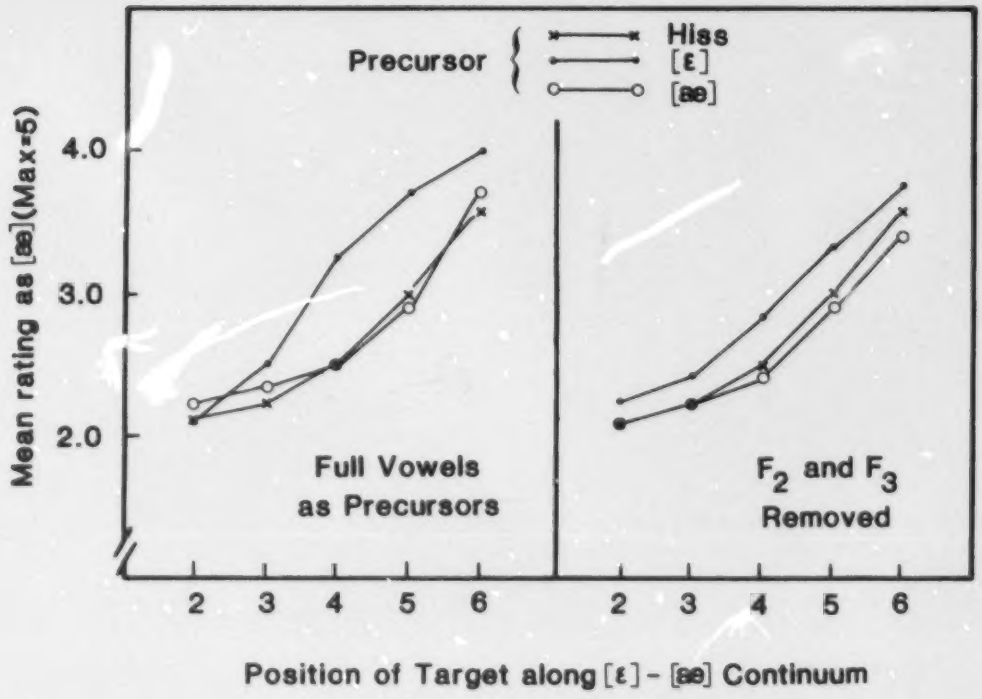


Figure 3. The same as Figure 1 except the data are for an /ɛ - æ/ continuum (Experiment 2).

and for the single formants, (2) this contrast was larger for the full vowels, however, than for the single formants, (3) there was no trace of contrast from the /æ/ direction, and (4) there was again a tendency for contrast to "bulge" in the most phonetically ambiguous region of the target items when the full vowels were precursors, but not when the single formants were used.

These observations were confirmed by several analyses of variance on mean ratings (toward /æ/). In the first of these, all five precursor conditions were crossed with the five target vowels. Both main effects and the interaction were statistically significant; for conditions, $F(4,116) = 13.32$, $p < .01$; for target vowels, $F(4,116) = 43.46$, $p < .01$; and for the interaction, $F(16,464) = 4.49$, $p < .01$. In the next analysis, the hiss condition was dropped, leaving a two-by-two design with respect to the precursor (full vowels versus F1 only X direction of contrast—from /ɛ/ versus from /æ/). The five target vowels were compared in the third factor. There were statistically significant main effects of whether the full or altered vowels were used as precursors, $F(1,29) = 6.33$, $p < .01$; of the direction of contrast, $F(1,29) = 47.12$, $p < .01$; and of the placement on the continuum of the target, $F(4,116) = 43.41$. Given the uniform results of precursors from the /æ/ direction, the first of these main effects means the full vowels produced more contrast than the first formants alone. The reliable interaction between direction of contrast and the five target vowels, $F(4,116) = 7.81$, $p < .01$, indicates that with the full and altered vowels combined, the interior (ambiguous) vowels were more affected than those closer to the endpoints. The three-way interaction between precursor type (full versus altered), direction of contrast, and vowel, was statistically significant here, as in the previous experiment, $F(4,116) = 3.96$, $p < .01$. This interaction is the most direct verification of the "bulging" in contrast effects for the full vowels but not for the altered ones.

The main empirical goal of this article is the establishment of single-formant contrast. Therefore, since single-formant contrast was smaller than full vowel contrast in this experiment, one additional analysis of variance was performed, including only two precursor conditions, the hiss control and the single-formant alteration of /ɛ/. In this analysis, the main effect associated with this comparison was reliable at the .01 level of confidence, $F(1,29) = 8.68$. The position of the target vowel was of course also a reliable source of variation, $F(4,116) = 41.29$, $p < .01$, but the interaction was less than 1.00. Thus, both experiments require the conclusion that single-formant contrast can occur on a vowel continuum even when these precursors do not resemble phonetically the vowels targeted for identification.

General Discussion

The two studies are so very consistent in most respects, we should deal with the single major discrepancy first: In Experiment 1, the total amount of contrast was no larger for the full vowels than for the single-formant precursors, but in Experiment 2, contrast was larger for the full vowels. One difference in the target stimuli used in the experiment may be critical here. In Experiment 2, there was variation between /ɛ/ and /æ/ in both of the first two formants, whereas in Experiment 1, only the first formant varied along the continuum. Methodologically, this discrepancy appears at first inexcusable but in fact was desirable to get "prototype" exemplars of both phonetic categories (see Peterson & Barney, 1952). The consequence is that in Experiment 2, an ambiguous item on the /ɛ/-/æ/ continuum was receiving potential contrastive influences from both F1 and F2 in the case of the full-precursors, but only from F1 in Experiment 1. It will be straightforward to untangle

these factors in future research should anyone ever be interested in whether single-formant contrast is numerically equal to regular contrast along continua varying in only one formant.

In both experiments, the full vowels produced markedly more contrast for the more ambiguous tokens than for those from near the endpoints (the "bulge"), whereas the altered, single-formant vowels produced relatively uniform contrast across the entire target continuum. We are intrigued by this result, but can offer little guidance in its interpretation. An obvious possibility for the locus of an interpretation is in some phonetic process. The identification targets that receive the most contrastive influence from the full vowels are those that are most ambiguous phonetically. Correspondingly, what distinguishes the full vowels from their single-formant variants is that they carry clear phonetic information about the relevant continuum. Somehow the richness of the phonetic information in the precursor could combine judgmentally with the precarious initial classification of the ambiguous target. But such a simple appeal to judgmental bias will obviously not work, because there is no corresponding selective effect of the putative phonetic process when contrast is measured from the "wrong" direction (that is, from either /ɪ/ or from /æ/).

No contrast was obtained in either study when one continuum endpoint was used and abundant contrast was obtained when the other was used. What is the principle responsible for these asymmetries? With data on only two continua, we would be foolish to propose a general hypothesis. We are now following several theoretical possibilities experimentally. The main burden of this paper is not the asymmetry in contrast observed here and elsewhere, however. In both experiments, we found unambiguous evidence that single-formant precursors affect ambiguous vowel identification. This much was predicted by Crowder's (1978, 1982, 1983) theory. There may well be other theories that predict single-formant vowel contrast and so we shall not stress the confirmation of these results for that one particular prediction. More to the point, the sort of theory that assigns contrast in this situation to specific, sensory processes is advanced by single-formant contrast. If subjects were trying to "balance out" their use of the response categories between their internal naming of the precursor and their explicit rating of the target, the single-formant precursors should have been much less effective.

One caveat needs to be added: Removal of F2 and F3 in Experiment 1 made /i/ sound like /u/ and /ɪ/ sound like either /u/ or /o/. Now /i/ and /u/ share the feature of being high vowels, whereas /ɪ/ and /o/ are both articulated with the tongue in a lower position. If subjects in Experiment 1 heard the altered vowels as an /u/-/o/ contrast (14 of the 25 subjects in the preliminary identification test did) and if they were somehow sensitive to the high-low feature, it is possible that they applied a judgmental bias with respect to that feature. That is, if they heard what they thought was /u/ as the first member of the pair, followed by an ambiguous token between /i/ and /ɪ/, they might be biased to pick the lower tongue-position alternative, /ɪ/. A similar argument can be applied to /ɛ/ and /æ/ in Experiment 2, which might engage both the high-low and front-back dimensions. We cannot dismiss this possibility on the basis of the present experiments. However, this sort of process would encounter just as much difficulty with the asymmetry of contrast along the various vowel continua as do other notions. Also, in Experiment 1, where we have data on identification of the precursors, only about half of the subjects would have been expected to identify the single formant vowels as an /u/-/o/ contrast, if we believe the 14/25 ratio of the preliminary experiment. Furthermore, this alternative explanation requires considerable abstraction of

distinctive vowel features, which would be important if true, but remains highly speculative now. And finally, the F1 dimension in vowel space is very highly correlated with tongue height, so that the tongue-height dimension is just another level of discourse in which one can talk about F1 frequency.

References

Crowder, R. G. (1978). Mechanisms of auditory backward masking in the stimulus suffix effect. Psychological Review, 85, 502-524.

Crowder, R. G. (1981). The role of auditory memory in speech perception and discrimination. In T. Meyers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech. Amsterdam: North-Holland.

Crowder, R. G. (1982). Decay of auditory memory in vowel discrimination. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 153-162.

Crowder, R. G. (1983). The purity of auditory memory. Philosophical Transactions of the Royal Society, Section B, 302, 251-265.

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. Word, 8, 195-210.

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects in stop consonant identification. Journal of Experimental Psychology: Human Perception and Performance, 4, 599-609.

Diehl, R. L., Lang, M., & Parker, E. M. (1980). A further parallel between adaptation and contrast. Journal of Experimental Psychology: Human Perception and Performance, 6, 24-44.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. Language and Speech, 5, 171-189.

Gilchrist, A. (1977). Perceived lightness depends on perceived spatial arrangement. Science, 195, 185-187.

Lindsay, P. H., & Norman, D. (1977). Human information processing (2nd ed.). New York: Academic Press.

Parducci, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. II). New York: Academic Press.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the identification of vowels. Journal of the Acoustical Society of America, 24, 175-184.

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance, 5, 129-145.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. Perception & Psychophysics, 30, 309-314.

Sawusch, J. R., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. Journal of Experimental Psychology: Human Perception and Performance, 7, 408-421.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. (1980). Contextual effects in vowel perception II: Evidence for two processing mechanisms. Perception & Psychophysics, 27, 421-434.

INTEGRATION OF MELODY AND TEXT IN MEMORY FOR SONGS*

Mary Louise Serafine,+ Robert G. Crowder,++ and Bruno H. Repp

Abstract. Two experiments examined whether the memory representation for songs consists of independent or integrated components (melody and text). Subjects heard a serial presentation of excerpts from largely unfamiliar folk songs, followed by a recognition test. The test required subjects to recognize songs, melodies, or texts and consisted of five types of items: (a) exact songs heard in the presentation; (b) new songs; (c) old tunes with new words; (d) new tunes with old words; and (e) old tunes with old words of a different song from the same presentation ("mismatch songs"). Experiment 1 supported the integration hypothesis: Subjects' recognition of components was higher in exact songs (a) than in songs with familiar but mismatched components (e). Melody recognition, in particular, was near chance unless the original words were present. Experiment 2 showed that this integration of melody and text occurred also across different performance renditions of a song and that it could not be eliminated by voluntary attention to the melody.

Introduction

Song is a universal artform that consists of two seemingly separate components, melody and text. In practice a song may derive from a pre-composed melody to which words are added, or from a pre-existing text later set to music. In fact a song may be the work of two artists, a composer and a poet or librettist. Yet the relationship between melody and text raises interesting questions in the domains of both aesthetics and cognitive psychology.

One of the aesthetic issues is how the artform should be defined: whether it is simply a pairing of independent components or an integral whole that transcends its parts. This issue has implications for the analysis of songs from a music-theoretic viewpoint--for example, whether the components can be considered or analyzed separately.

A parallel issue can be raised from a cognitive viewpoint: To what degree are melody and text independent or integrated in perception and memory? While there is substantial literature both on linguistic memory and on musical

---

+Department of Psychology, Yale University.
++Also Department of Psychology, Yale University.
Acknowledgment. This research was supported by NSF Grant 8005838 to Robert G. Crowder and by NICHD Grant HD-01994 to Haskins Laboratories. Mary Louise Serafine is now at the Department of Psychology, Vassar College, Poughkeepsie, New York 12601.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

memory (Deutsch, 1969; Dowling, 1973, 1978; Dowling & Fujitani, 1971), research thus far does not indicate how a hybrid form such as songs might be represented in memory. Indeed, research on hemispheric differentiation, especially that which suggests left-hemisphere dominance for language and right-hemisphere dominance for music (e.g., Best, Hoffman, & Glanville, 1982; Kimura, 1967), leaves entirely open how melody and text in songs might be processed.

Our interest in this issue was generated by informal observations suggesting that, in memory for songs, melody and text form an integrated unit, such that people find it difficult to separate the two components. For example, if asked to recite the words of their national anthem, many people would have to sing the song, or at least rehearse it subvocally, in order to generate the words. Also, people may not immediately recognize that two different songs have the same melody if their texts are different. "Twinkle, Twinkle, Little Star" and "Baa, Baa, Black Sheep" are a case in point, where identical melodies are part of what are considered entirely different songs. Yeston (1975) provides the example of the well-known theme of the Mozart C major piano sonata (K. 545), which (with slight changes in rhythm) is rarely recognized as the melody of "Hey There, You With the Stars in Your Eyes." Finally, the first author has found instances of profound melody/text integration in informal experiments with a young child. In these experiments a two-year-old, who could repeatedly and accurately perform a large body of songs, was nevertheless incapable of singing the melodies on the syllable "la" without the words. Instead, she simply spoke the syllable in rhythm. Similarly, she was either unwilling or unable to repeat the words without the melody.

These examples argue for some form of integration of melody and text in memory for songs, although it is also true that adults, at least, can voluntarily separate a melody from its text and vice versa in singing and recognition. Thus in theory the memory representation for songs might consist of: (1) independent components, (2) integrated components, or (3) a non-decomposable whole (an extreme form of integration). If melody and text were stored as independent components, we would expect that memory for songs could be predicted by the independent probabilities of memory for melody and memory for text. On the other hand, if the components were integrated, we would expect that memory for one component facilitates memory for the other. Finally, if songs were stored as non-decomposable wholes, we would expect that melodies cannot be recognized as familiar when their words are different, and vice versa. This last hypothesis is clearly false in many situations: Words are easy to recognize in new contexts, and most people can probably recognize a tune when the words are different if the tune is pointed out to them. (Musicians and experienced listeners can often do so in any case.) Nevertheless, it is worth investigating the degree to which a wholistic representation may characterize novel songs, when attention is not explicitly drawn to tune-similarity.

Note that the issue of integration can be distinguished, at least conceptually, from two related issues: compatibility and association. Melodies and texts are often compatible rhythmically in that higher pitches, longer durations, and musical-metric stresses tend to occur on accented syllables. Similarly, melodies and texts may be compatible "semantically" in that the tempo and musical mood seem to fit the meaning of the words. However, it is possible that a cognitive form of integration occurs irrespective of the

compatibility of components. Indeed, whether compatibility is necessary or sufficient for integration is an empirical question not under consideration in the present experiments.

Also, integration of melody and text in song can be distinguished from association as mere knowledge of co-occurrence. That melody and text co-occur is undeniable. Yet association may occur without integration. Indeed, it is possible to imagine other co-occurring events (e.g., speech and background music) that do not give rise to integration.

The purpose of these experiments, then, was to investigate the degree to which melody and text are independent, integrated, or nonseparable ("wholistic") in memory for songs.

Experiment 1

Subjects heard 24 consecutive excerpts from folk songs, followed by a 20-item recognition test. The test items were of two types: (1) excerpts that had been heard in the presentation ("old songs") and (2) excerpts that had not been heard in the presentation ("new songs"). Further, new songs were of four types: (a) new tune with new words; (b) old tune with new words; (c) new tune with old words; and (d) old tune with old words that had been sung to a different tune in the original presentation ("tune and words mismatched"). In the remainder of the paper, the terms "tune" and "words," as used with subjects, are interchangeable with "melody" and "text."

The main prediction was that, if subjects integrate melody and text in memory, they should recognize previously heard melodies or texts more accurately when they are paired with their original companion (text or melody) than when they are paired with a different companion. On the other hand, if melody and text are stored as independent components, then subjects should recognize previously heard melodies or texts equally well, whether paired with the same or with a different companion. Finally, if songs are stored as wholistic units, then subjects should not be able to recognize melodies (or texts) at all, except when they are paired with their original companion.

Method

Materials. Songs that we considered unfamiliar to the average listener were drawn from a collection of indigenous American folk songs compiled by Erdei (1974). Twenty pairs of excerpts with interchangeable melodies and texts were chosen, each excerpt consisting of the opening two to four measures of a song. (See list in appendix.) Thus each pair of excerpts yielded four different songs, a total of 80. Figure 1 shows a sample pair of interchangeable melodies and texts. Examples of the five test-item types are shown in Figure 2.

In some cases minor alterations were made to the original melody or text to ensure a rhythmic fit with its companion. (See appendix.) For example, "across" from one original text was changed to "'cross" in our experiments (Figure 2, test item a). However, in all cases the texts and melodies were identical across parallel presentation and test versions of a song.

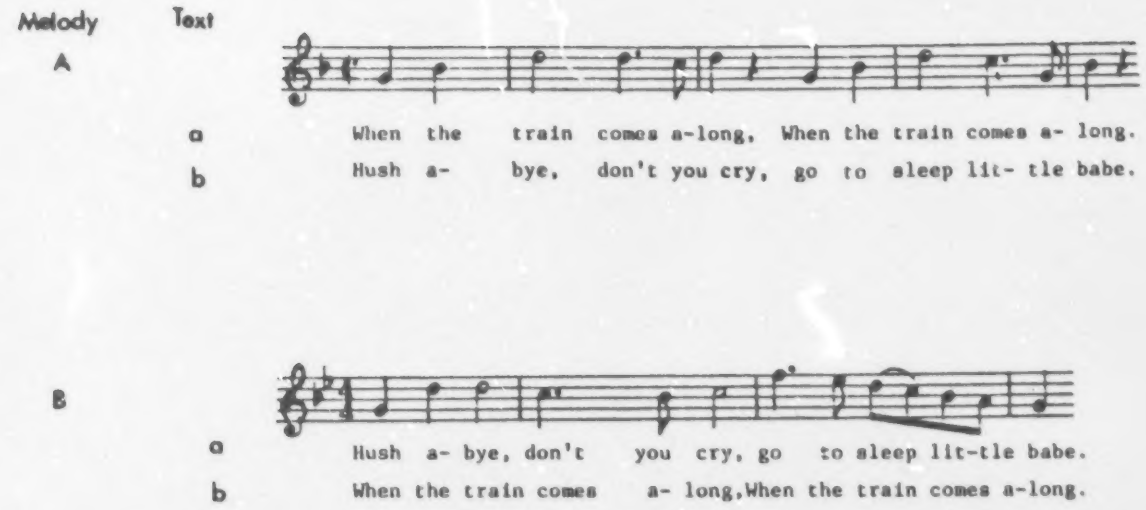

Melody
Text
A
a
When the train comes a-long, When the train comes a- long.
b
Hush a- bye, don't you cry, go to sleep lit- tle babe.
B
a
Hush a- bye, don't you cry, go to sleep lit-tle babe.
b
When the train comes a- long,When the train comes a-long.

Figure 1. Sample pair of songs with interchangeable texts. (Aa and Bb denote original songs; Ab and Ba denote derivatives.)

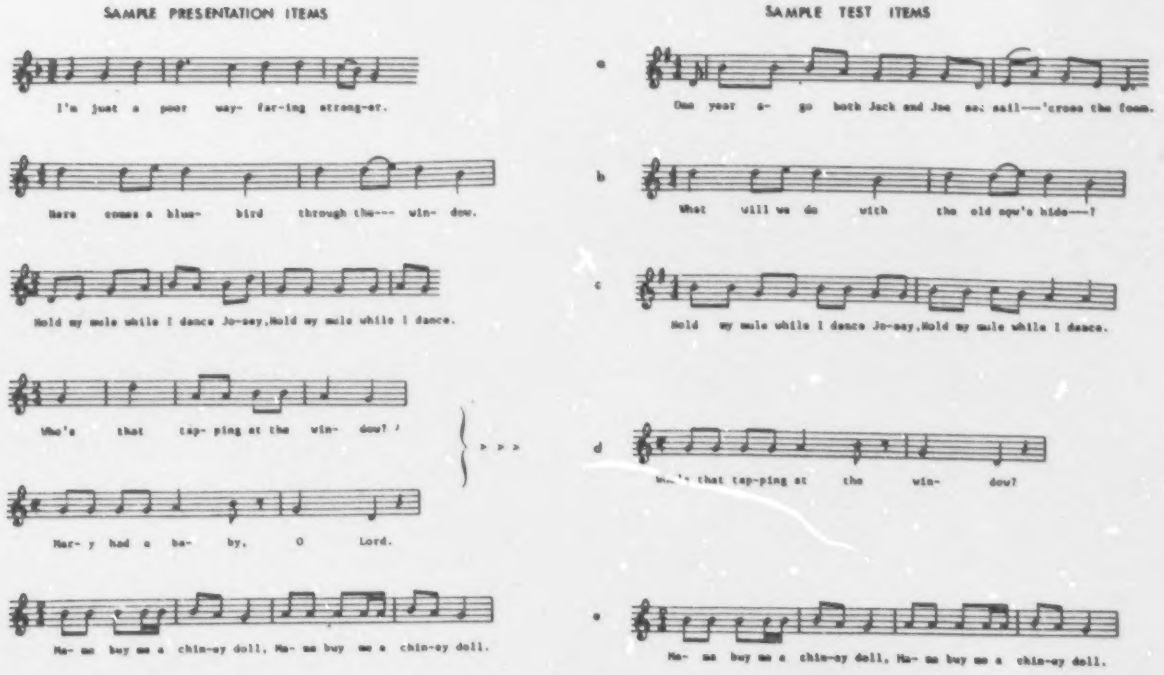

SAMPLE PRESENTATION ITEMS
I'm just a poor way- far-ing strang-er.
Here comes a blue- bird through the--- win- dow.
Hold my mule while I dance Jo-sey,Hold my mule while I dance.
Who's that tap- ping at the win- dow? /
Mar- y had a ba- by, O Lord.
Ma- ma buy me a chin-ey doll, Ma- ma buy me a chin-ey doll.
SAMPLE TEST ITEMS
a
One year a- go both Jack and Joe set sail---'cross the foam.
b
What will we do with the old cow's hide---?
c
Hold my mule while I dance Jo-sey,Hold my mule while I dance.
d
W..'s that tap-ping at the win- dow?
e
Ma- ma buy me a chin-ey doll, Ma- ma buy me a chin-ey doll.

Figure 2. Sample presentation and test items. (a: new tune, new words; b: old tune, new words; c: new tune, old words; d: old tune, old words—mismatched; e: old song)

The excerpts, sung by a tenor with vocal training, were recorded on tape. They were sung as notated, except transposed down a fifth (or twelfth) to the tenor range, and at a tempo of one beat per second (MM: ♩ = 60). The excerpts varied in key and mode, but were notated with G as the tonic in each case. The tapes were recorded with a 5-sec interval of silence between presentation items and a 10-sec response interval after each test item.

Design. From the bank of 80 song excerpts, five parallel sets of presentation and test sequences were constructed. Each set was administered to a different group of subjects.

In the presentation sequences (24 items), half the excerpts were tunes with their original words (type Aa or Bb in Figure 1), and half were tunes with words borrowed from their companion song (type Ab or Ba in Figure 1).

In the test sequences (20 items), each of the five test item types (a through e in Figure 2) occurred four times. Moreover, across the five subject groups presentation items were paired with each of the five possible test item types, following a Latin square design. For example, Table 1 shows the generation of possible presentation and test items from two of the song pairs.

Table 1

Presentation and Test Items From Sample Song Pairs

SUBJECT GROUP	I		II		III		IV		V	
SONG PAIR[1]	Pres./Test		Pres./Test		Pres./Test		Pres./Test		Pres./Test	
		c		a		b		d		e
[Aa and Bb]$_X$	Aa	Ba	Aa	Bb	Aa	Ab	Aa Bb	Ab	Aa	Aa
		b		c		e		a		d
[Aa and Bb]$_Y$	Ab	Aa	Ab	Bb	Ab	Ab	Ab	Ba	Ab Ba	Aa

a = new tune, new words
b = old tune, new words
c = new tune, old words
d = old tune, old words, mismatched
e = old song

[1]Two examples (X, Y) from 20 song pairs.

As shown in Table 1, a mismatch test item (type d: old words paired with old tune of a different song) required two presentation excerpts. Whenever two such items were required in the presentation, they immediately followed each other on the tape. Thus each presentation sequence required 4 pairs of songs for the mismatch test items, plus 16 songs for the other test item types (total of 24 songs).

In all presentation and test sequences, the excerpts were generated successively from Song Pairs 1 through 20, in the order listed in the appendix. Thus, the interval between each presentation item and its corresponding test item was roughly constant.

Procedure. Subjects were tested in small groups in a quiet room, except for one large group that was tested in a classroom. Presentation and test tapes were played back over loudspeakers. Subjects were instructed to listen carefully to a presentation of 24 excerpts from simple folk songs and told that their memory for them would be tested later. After the presentation they were asked whether any of the excerpts were familiar, and if so, to estimate their number on the answer sheet. Following that, the answer sheet was explained and the test sequence was presented. About five minutes elapsed between the presentation and test sequences.

For each test item, subjects were asked to indicate on the answer sheet whether they had "heard that exact excerpt before," and if not, whether they had heard either the tune or the words. In advance of the test, subjects were given an explanation of the term "tune" (melody) and a description of the five types of items they could expect on the test (types a through e). Thus, they were prepared for the test of recognizing tune, words, or exact song, but they did not have knowledge of this requirement prior to the presentation.

Subjects. Subjects were 32 undergraduate students with varying degrees of musical training. The first 16 subjects, who were tested together in a classroom environment, necessarily were all assigned to one particular presentation/test condition. Sixteen additional subjects were divided among the remaining four conditions.

Results and Discussion

Subjects' post-presentation estimates of the number of songs that seemed familiar averaged 1.4 (out of 24 presentation items, 12 of which were original folk songs). This result confirmed the relative unfamiliarity of the materials.

For the discussion of recognition scores, we adopt this terminology: If subjects indicated that they recognized an exact test excerpt as one that had been heard in the presentation, this is called an "old song" response. Similarly, if they indicated recognition of just the melody or the text, this is called an "old tune" or "old words" response, respectively.

Recognition of old songs. Table 2 lists the mean proportion of "old song" responses made to the five types of test items. Subjects correctly recognized old songs 85% of the time, a surprisingly high recognition rate, given that the presentation excerpts had been heard only once. Incorrect responses were lowest (.07 and .06) whenever new words were heard, and highest (.39) for "mismatched" tune and words, where both components had been heard originally in different presentation songs.

Table 2

Mean Proportion of "Old Song" Responses (Exp. 1)

		Words		
		New	Old	Mean
New Songs				
Tune	New	.07	.25	.16
	Old	.06	.39	.22
	Mean	.06	.32	
Old Songs				.85

Table 3

Mean Proportions of "Old Tune" and "Old Words" Responses (Exp. 1)

		"Old tune" responses			"Old words" responses			
		Words				Words		
		New	Old	Mean		New	Old	Mean
New Songs								
Tune	New	.44	.40	.42	New	.10	.78	.44
	Old	.53	.63	.58	Old	.13	.85	.49
	Mean	.48	.52		Mean	.12	.82	
Old Songs				.92				.92

It might be argued that this high false alarm rate for mismatched old tune and words (.39) indicates some measure of independent storage for song components. To some degree, subjects erroneously thought they recognized the mismatch songs, apparently because the components were familiar, though never paired in the original presentation. Note, however, that this effect was largely due to a high false-alarm rate for all items containing old words, which probably reflects the fact (discussed below) that words were much easier to remember than tunes. Nevertheless, this false alarm rate is far below the hit rate for original old songs, which indicates that subjects were more likely to retain the association of a presented melody with its presented text than to retain the components independently.

This is necessary but not sufficient evidence that subjects integrated melody and text. To address the issue of integration, we must examine melody and text recognition separately, and determine whether these components were more accurately recognized in old songs than in any type of new song.

Recognition of components. Table 3 shows the mean proportion of responses to the questions regarding recognition of old songs, tunes, and words. In this table "old song" responses are included both in "old tune" and in "old words" responses, for a response of "old song" indicates that subjects recognized both the tune and the words.

The main hypothesis was that, if melody and text are integrated in memory, old tunes should be recognized more accurately in old songs than in mismatch songs (or songs with new words). Similarly, old words should be recognized more accurately in old songs than in mismatch songs (or songs with a new tune).

Consider first the "old tune" responses. Old tunes were recognized more accurately in old songs (.92) than in mismatch songs (.63) or songs with new words (.53). The advantage for old songs over mismatch songs was highly significant across subjects, $t(32) = 5.27$, $p < .001$, and across test items, $F(1,18) = 16.20$, $p < .001$. The advantage was equally large for old songs presented in their original folk song version and for old songs constructed by recombining the melodies and texts of different original folk songs, $F(1,18) = 0.05$.

Consider now the "old words" responses. Words were recognized more accurately in old songs (.92) than in mismatch songs (.85) or songs with new tune (.78). Because of ceiling effects, the advantage for old songs over mismatch songs fell just short of significance across subjects, $t(32) = 2.00$, $p < .06$, but it was significant across items, $F(1,18) = 6.35$, $p < .02$. Once again, it did not matter whether or not the old song was a real folk song, $F(1,18) = 0.18$.

These results suggest that melody and text are integrated in memory to a considerable degree. One component is recognized better in the context of the other, original component, than in some new context. The advantage for original contexts (old songs) holds even over new contexts in which the components are just as familiar (mismatch songs). The deciding factor seems to be not whether the components are familiar, but rather whether they had been paired in the initial perception. Thus melody and text appear not to be stored independently; the components are stored in some integrated fashion.

Responses to new songs. The data from Experiment 1 allow for a further clarification of the integration effect. In the remaining discussion we consider "old tune" and "old words" responses to new songs only. Two issues are of interest here. First, were tunes and words recognized better than chance in these new contexts? A strong form of the integration hypothesis--that is, a "wholistic" conception--would predict that tunes and words cannot be recognized at all outside of their original contexts (old songs). Second, aside from the integration effect described above, there might also be a "contamination" effect from companion components at the recall (not storage) stage. That is, were tune and word judgments (whether correct or incorrect) influenced by the familiarity of the other component?

To examine the above issues, separate 2 X 2 ANOVAs were performed on "old tune" and "old words" responses, both across subjects and across items, with the factors of tune (old vs. new) and words (old vs. new) whose combination represents the four types of "new song" test items. With regard to tunes, Table 3 shows a mean hit rate of .58 and a mean false alarm rate of .42, which represents rather poor performance. The difference between hits and false alarms was significant across subjects, $F(1,30) = 9.61$, $p < .01$, but not across items, $F(1,18) = 2.99$, $p < .11$. Thus, tune recognition in new songs was near chance. The recognition score for words was much higher: a mean hit rate of .82 versus a mean false alarm rate of only .12. This difference was highly significant, of course.

Thus the strong form of the integration argument--a "wholistic" conception--does not hold up to test here. Certainly, texts were recognized better than chance in new contexts, and there seemed to be some minimal memory for tunes as well, indicating some degree of independent storage of components. As discussed earlier, components are more accurately recognized in original contexts (old songs), but they may also be recognized to some degree in new contexts.

The second issue concerns the influence of one component's familiarity on judgments of the other component--a "contamination" effect. With respect to tunes, Table 3 reveals that subjects responded "old tune" more frequently when the words were old (mean of .52) than when the words were new (mean of .48). This small effect was significant across subjects, $F(1,30) = 5.91$, $p < .05$, but not across items, $F(1,18) = 1.35$. With respect to words, subjects responded "old words" somewhat more frequently when the tune was old (mean of .49) than when it was new (mean of .44). This effect was also significant across subjects, $F(1,30) = 5.52$, $p < .05$, but not across items, $F(1,18) = .05$.[1]

In summary, Experiment 1 yielded the following results. The main finding was that recognition of one component (melody or text) was facilitated by the simultaneous presence of the other, original component (in old songs). This effect argues for an integrated representation of melody and text in memory for songs. In addition, we found that recognition memory for old songs was excellent, even after a single presentation. Our casual observation was that this excellent performance was accompanied by rather low confidence: Many subjects felt they were just guessing.

However, there is evidence that tunes, and especially words, can be recognized to some degree when paired with new components. While this does not contradict the integration hypothesis, it does indicate some measure of

separation of components and argues against the stronger "wholistic" conception of melody/text relations.

Experiment 2

The integration effect in Experiment 1 leaves open at least three questions. First, perhaps the effect was induced by the requirements of a song (rather than melody or text) recognition task. The testing procedure in Experiment 1 required primarily "old song" recognition, and only conditionally tune and word recognition. Thus tune and word recognition scores were based in large part on correct "old song" responses. It remains to be determined whether subjects would recognize tunes or words more accurately in old than in new contexts if they were asked to judge only these components. In other words, if "old song" responses were not permitted, would there still be an advantage for tune or word recognition in old songs?

A second issue concerns the extent to which the integration effect is sensitive to subjects' strategies at the presentation stage. In the first experiment, subjects listened to the presentation songs with the knowledge that their memory for songs would be tested. Perhaps this instruction engendered a global, integrated memory for melody and text at the presentation stage. What remains to be determined is whether this integration is optional or obligatory. In other words, would the integration effect still hold if subjects were given the instruction to listen analytically? For example, if subjects were told at the presentation stage that their memory for tunes would be tested, would they be able to ignore the words?

A third question concerns the generality of the integration effect. In the first experiment, the presentation and test tapes were recorded by the same performer. Thus vocal inflection, timbre, and other variables in the performance of melodies and texts would be similar across presentation and test songs. It remains to be determined whether the integration effect is sufficiently abstract to hold across different performance renditions of a song. In other words, would the integration effect hold even for a recognition test in which the items are sung by a different performer? Moreover, a possible danger to avoid is that old song recognition might be an artifact of the acoustical identity of old songs across the presentation and test tapes. Any physical identity, even an accidental or musically irrelevant one, could have contributed to the old song recognitions in Experiment 1. If the integration effect were found to hold across different performers, it would prove to be abstract as well as unattributable to the acoustical identity of old songs.

Experiment 2 was designed to address these issues. Specifically, Experiment 2 sought to determine (1) whether the integration effect would hold in a melody-only rather than song recognition task; (2) whether it would hold even in the face of instructions to listen analytically--that is, to tunes only--at the presentation stage; and (3) whether it would hold across different performers (and performance renditions) of the presentation and test songs.

Method

Materials. The materials were the same as in Experiment 1, except that the five sets of presentation and test sequences were recorded a second time, this time by a female vocalist in the alto range, a perfect fifth higher than

the male tenor recordings. While the same general guidelines were followed as to tempo and and other notated musical factors, no attempt was made to imitate the tenor's performance renditions.

Design. The recordings by male and female vocalists allowed for four combinations of male and female presentation and test sequences (M/M; M/F; F/M; F/F). These four conditions were further subdivided into two instruction conditions. The resulting eight conditions were applied across the five sets of presentation/test sequences. This resulted in 40 conditions.

Procedure. The procedure was similar to that of Experiment 1, with the following differences: Half the subjects received the same instructions as in Experiment 1. The other half received "analytic" instructions: "Listen carefully to these songs and your memory for the tune or melody only--that is, just the musical portion--will be tested later. You can ignore the words because you will not be tested on these." At the test stage, all subjects made a written response to the question, "Did you hear this exact melody before?" for each item.

Subjects. Subjects were 48 undergraduate students of varying musical backgrounds. Each of 40 conditions contained one subject. Eight additional subjects were assigned to the first set of the presentation/test tapes, distributed across the eight conditions of performance rendition and instruction.

Results and Discussion

Subjects in Experiment 2 found the folk song materials as unfamiliar as had subjects in Experiment 1. After the presentation of 24 songs, subjects reported a mean of 1.2 familiar songs.

Recognition of tunes. Table 4 compares tune recognition in mismatched new songs and in old songs for two conditions of performance rendition (same voice vs. different voice) and two conditions of instruction (general vs. analytic). These data were analyzed in a three-way ANOVA across subjects. For reasons having to do with the design of the experiment, the performance and instruction factors were not included in the ANOVA across items.

The results confirmed the integration effect found in Experiment 1. That is, even in this tune recognition task, subjects recognized tunes more accurately in old songs (mean of .84 across all conditions) than in mismatch songs (mean of .64), $F(1,40) = 17.19$, $p < .001$, across subjects and $F(1,18) = 13.42$, $p < .002$, across items. Moreover, the integration effect was maintained even for the analytic condition, where subjects were told to pay attention only to tunes at the presentation stage. There was no significant main effect for instructions or any interaction in this analysis.

Further, the integration effect held to a considerable degree even across different performance renditions. Although Table 4 suggests that the advantage for old songs was reduced in the different-performer condition, this interaction of test item type and performance rendition was not significant, $F(1,40) = 3.86$, $p < .10$. Finally, as in Experiment 1, whether or not an old song was a real folk song made no difference, $F(1,18) = 1.31$.

Table 4

Mean Proportion of "Old Tune" Responses (Exp. 2)

Performance:	Same		Different	
Instructions:	General	Analytic	General	Analytic
New songs (mismatch)	.60	.69	.69	.60
Old songs	.94	.94	.73	.77

Table 5

Mean Proportion of "Old Tune" Responses: New Songs (Exp. 2)

		General Instructions				Analytic Instructions		
		Words				Words		
		New	Old	Mean		New	Old	Mean
Tune	New	.20	.57	.38	New	.36	.50	.43
	Old	.23	.64	.44	Old	.35	.64	.50
	Mean	.21	.61		Mean	.35	.57	

We thus conclude that the integration effect is robust. Melody and text appear to be integrated in memory, even in the face of attempts to focus on or separate the melody at the presentation stage, and even when the performer is different at the recognition stage.

Responses to new songs. It remains to be determined how the effects of instruction and performance rendition influenced (1) the accuracy of tune recognition in new songs, and (2) the "contamination" effect of words on "old tune" responses. The relevant data are shown in Table 5, separately for the two instruction conditions but averaged over performance conditions, which showed no effect here.

The data for new songs were analyzed in an ANOVA across subjects on "old tune" responses with the factors of tune (old vs. new), words (old vs. new), instruction condition (general vs. analytic), and performance rendition (same vs. different). In the ANOVA across items, only the first two factors were included.

As Table 5 shows, tune recognition in new songs was poor, even worse than in Experiment 1. The main effect for tunes was not significant in either analysis. Although it may appear that subjects had some success in recognizing tunes when the words were old (compare hits with false alarms in "old words" columns), in fact the tune x words interaction was not significant. Thus subjects did not recognize tunes better than chance in new song contexts. Moreover, tune recognition was equally poor regardless of instructions or performance renditions.

However, there was a highly significant main effect for words, with subjects giving many more "old tune" responses when the words were old (mean of .59) than when the words were new (mean of .28), $F(1,40) = 50.01$, $p < .0001$, across subjects, and $F(1,18) = 37.58$ $p < .0001$, across items. In addition, this effect interacted with instructions, $F(1,40) = 4.15$, $p < .05$, in that it was less pronounced in the analytic instruction condition.

In summary, Experiment 2 showed that the integration effect for memory of original melody and text is both obligatory and abstract. Analytic instructions did not reduce the integration effect; subjects were unable to ignore the words in storing melodies at the presentation stage. Moreover, the integration effect is generalizable across different performance renditions in the presentation and test stages.

That tunes were recognized so poorly in new contexts would seem to argue for an even stronger form of the integration hypothesis—a "wholistic" conception of melody/text relations in memory. Even instructions to listen analytically did not improve tune recognition. While it seems possible that there is an asymmetry in memory integration, such that tunes are more dependent on the words than vice versa, our findings may simply reflect the fact that the tunes were much harder to remember than the words. This may be an artifact of the folk song genre, since the melodies were in many ways similar (small range; G tonic; homogeneous rhythm; mostly step-wise melodic motion), but the texts were very different from each other. Moreover, texts could be recognized by a single salient word (e.g., "Babylon" or "turkey"), but the tunes had no such advantage. Ultimately, the question of which component is more memorable boils down to the nature of the materials. We might imagine a reversal of the memory advantage for words if we had selected texts that were very similar to

each other, and melodies that were widely discrepant. But in the case of the folk songs used in our experiments, a natural asymmetry exists in the salience of texts and melodies.

General Discussion

We conclude from these experiments that melody and text are integrated in memory to a considerable degree. We found that the familiarity of old tunes and words (when mismatched) was an insufficient predictor of the superior recognition for original old songs. Moreover, we found no evidence that subjects can voluntarily reduce the degree of integration of melody and text. Indeed, what was surprising was not only the size of the integration effect, but that subjects seemed to be unaware of it. Thus melody and text appear not to be stored as independent components. On the other hand, a stronger or "wholistic" form of integration appears to be untenable, at least as far as the text is concerned. Our results leave open the possibility that, under certain conditions, the melody may be completely integrated with the text (but not vice versa).

In addition to this integration, there appears to be a reciprocal "contamination" in familiarity judgments of melodies and texts. This effect may be voluntarily reduced, though not entirely removed. The effect itself may be an artifact of the selected materials, and it may depend on other factors that are not clear at present; we have no explanation for the difference between Experiments 1 and 2 in the magnitude of the influence of word familiarity on tune judgments.

One question that is left unresolved by the present experiments is the degree to which tune recognition in old songs may have been due to subtle changes imposed on a melody by the specific texts employed. Two possibilities are a semantic effect and a prosodic effect. For example, specific semantic connotations may become associated with a melody when it is heard in connection with a text about animals, cobblers, lullabies, dancing, and so forth. These connotations may facilitate tune recognition in old songs or hinder recognition when the text is different. To take an extreme example (Figure 2, item b), it may be difficult to recognize a melody originally heard in connection with a bluebird coming through a window, when that melody is later heard in connection with a "old sow's hide."

An alternative hypothesis is that different texts impose prosodic or submelodic variations on melodies. A change in text results in a drastic change in the segmental structure of the words, which may have modified to some extent what was nominally the same melody. For example, different patterns of consonants, vowels, stresses, and voicing may influence the onset and decay characteristics of tones and the precise degree of stress given to them. Thus similarity of submelodic structure may have facilitated tune recognition, even across different performance renditions, although it can hardly account for the whole old song advantage.

We note here a natural asymmetry in the relation between (audible) melody and text: While a tune can exist perfectly well without any words (when played on a musical instrument, for example), words always have some kind of "tune," if only the nonmusical one provided by the prosody of spoken language. In the context of a song, the musical tune in large measure takes over the function of prosody and thus becomes an aspect of the suprasegmental proper-

ties of the words. Viewed in this way, it is quite conceivable that memory for tunes is more dependent on memory for words than vice versa; certainly, outside the realm of music the prosody of speech is remembered, if at all, only as an aspect of the words by which it is carried. We hope to investigate this interesting parallel between speech and music in future experiments.

References

Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. Perception & Psychophysics, 31, 75-85.

Deutsch, D. (1969). Music recognition. Psychological Review, 76, 300-307.

Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. Perception & Psychophysics, 4, 37-40.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. Psychological Review, 85, 341-354.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. Journal of the Acoustical Society of America, 49, 524-531.

Erdei, P. (1974). 150 American folksongs to sing, read, and play. New York: Boosey and Hawkes.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. Cortex, 3, 163-178.

Yeston, M. (1975). Rubato and the middleground. Journal of Music Theory, 19, 286-301.

Footnote

[1]We noted that both of these "contamination" effects were exhibited only by one group of subjects (the large group assigned to Condition I) but not by the other groups. We have no explanation for this difference. With respect to all other effects of interest, the subject groups gave equivalent results.

APPENDIX

Pairs of Folk Song Excerpts with Interchangeable Texts.

All Folk Songs from Erdei (1974)

	Number/Title		Number/Title	
1	9:	Hunt the slipper	92:	Cape Cod Girls
2	12:	Let us chase the squirrel[1]	73:	Christ was born[1]
3	15:	Who's that tapping at the window?	82:	Mary had a baby
4	16:	How many miles to Babylon?[1,2]	120:	Nuts in May
5	21:	Poor little kitty puss!	80:	Turn the glasses over
6	22:	Down in the meadow	68:	The old woman and the pig
7	27:	Hush little baby	13:	Bye, bye baby
8	32:	Bluebird	55:	The old sow
9	38:	Ida Red[1,2]	39:	Mama, buy me a chiney doll
10	52:	Dear companion	88:	Wayfaring stranger
11	67:	I lost the farmer's dairy key	128:	Watch that lady
12	69:	Old turkey buzzard	72:	My good old man
13	78:	Hold my mule	102:	Needle's eye
14	99:	When the train comes along	132:	Hushabye[1,2]
15	103:	Housekeeping	147:	My old hen[1]
16	148:	I'm goin' home on a cloud	138:	The raggle taggle gypsies
17	110:	Give my love to Nell[1]	137:	Blow, boys, blow
18	122:	Cripple Creek	129:	The little dappled cow
19	142:	Goodbye girls, I'm going to Boston	144:	Cradle hymn
20	2:	The boatman	86:	The Derby ram

---

[1] Minor alteration was made in text
[2] Minor alteration was made in melody

THE EQUATION OF INFORMATION AND MEANING FROM THE PERSPECTIVES OF SITUATION SEMANTICS AND GIBSON'S ECOLOGICAL REALISM*

M. T. Turvey+ and Claudia Carello++

Introduction

As Barwise and Perry suggest, their theory of meaning is consistent on several fronts with Ecological Realism as it has been developed by the psychologist James J. Gibson. The most important convergence from our perspective is the shared conviction that meaning is neither in the brain--the residence openly preferred by orthodox psychologists--nor in some netherworld--a location intimated by Fregean semantics. Rather, meaning is contained in the system defined by the nested relations between the real properties of a living thing and the real properties of the environment with respect to which the living thing conducts its daily affairs.

How is this type of realist account of meaning supported? Both Gibson and Barwise and Perry have attempted to ground meaning in information. But both are extremely careful about the sense in which information is to be used. Gibson (1966) pointed out that information theory in the style of Shannon (1949) was not adequate to the demands of perceiving--obtaining information about activity-relevant properties of the environment. Whereas information for communication engineering is assumed to be finite and transmittable, information for perceptual systems is inexhaustible and noticeable (i.e., not carried, as through a channel) (Gibson, 1979). To characterize information as a quantifiable reduction in uncertainty does not require a consideration of meaning; to characterize information as the specification of the observer's environment demands it. Similarly, Barwise and Perry deny Dretske's (1981) assertion that meaning and information are dissociable. Instead, situation semantics and ecological psychology place what Barwise and Perry call "constraints on the structure of reality" at the heart of their attempts to consider meaning and information conjointly. That is to say, if an event, A, is linked systematically to another event, B, A is information about B; the linkage is meaningful.

It is in the nature of the linkage that the different emphases of the two approaches can be seen. Gibson (1954, 1966) identified three such linkages: convention, projection, and natural law. These underwrite the relationships between, for example, an automobile and its license, an automobile and its

---

*To appear in: Linguistics and Philosophy (special issue on Barwise, J., and Perry, J. Situations and attitudes)

+Also University of Connecticut.

++State University of New York at Binghamton.

Acknowledgment. The writing of this paper was supported in part by ONR Contract N000 14-83-C-0083

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

shadow, and a moving automobile and the optical flow pattern it generates, respectively. Examples of the last type--what Barwise and Perry refer to as information based on nomic structural constraints--are at the core of the Gibsonian program. An understanding of the information required for animals to control locomotion in a cluttered surround is considered propaedeutic to understanding information of the other types. The focus is on uncovering laws at the ecological scale (i.e., appropriate to a given animal-econiche system) (Turvey, Shaw, Reed, & Mace, 1981) that underlie information in the specificational sense (Reed, 1981; Turvey & Kugler, 1984, in press), captured as follows:

$$\text{Situation-type A} \quad \underset{\text{specifies}}{\overset{\text{generates by law}}{\rightleftharpoons}} \quad \text{Situation-type B} \qquad (1)$$

Information in the pictorial sense and information in the indicational sense (that central to linguistic meaning) can be schematized similarly:

$$\text{Situation-type C} \quad \underset{\text{depicts}}{\overset{\text{produces by projection}}{\rightleftharpoons}} \quad \text{Situation-type D,} \qquad (2)$$

and

$$\text{Situation-type E} \quad \underset{\text{indicates}}{\overset{\text{is conventionally linked to}}{\rightleftharpoons}} \quad \text{Situation-type F,} \qquad (3)$$

respectively. For Gibson, both of these are predicated on information in the specificational sense. A representational picture, for example, is a surface treated in such a way as to make available some of the same (formless and timeless) invariants that are available in the real scene (Gibson, 1979). In the same vein, the symbolic waggle dance of the bee indicates the location of a source of honey in the invariant pattern of dips and twists. In all cases, meaning is there to be discovered, whether the animal is immersed in a lawfully structured sea of energy, encounters an arrested array of persisting invariants, or confronts culturally determined conventions. That is to say, even if the constraints are at some remove from the animal-environment system, each new individual need not reinvent or recreate them. Rather, the systematicity of the relationships must be noticed. But the fundamentality of information in the specificational sense runs still deeper. In order for information in the indicational sense to be efficacious, information in the specificational sense already must be available. For example, in order for a stop sign to regulate the dynamics of traffic flow and, therefore, for its meaning to be realized, information specifying the retardation of forward motion and the time-to-contact with the place where velocity must go to zero, must be available.

We will pursue the notion of information in the specificational sense in the section that follows in an effort to support the arguments of Gibson and Barwise and Perry that meaning and information can be equated.

Information in the Specificational Sense and Situation-type Meaning

Consider a transparent medium (air or water) that is densely filled with light scattered by a substantial surface below. Now consider a point of observation that is moving in the medium rectilinearly relative to the ground. In order to define an optical field that flows relative to the point of observation, each point of the ambient light can be assigned a vector that is opposite that of the vector of the point of observation. If, for example, the point of observation is moving toward a point, then the optical field will flow outwards from that 'target point'. That is, there is a lawful relation of the type: forward rectilinear motion of a point of observation (F) ----> global optical outflow (O). (This is an instance of a more general law of ecological optics formulated as: a particular motion of a point of observation relative to a surround ----> a particular global transformation of the ambient optical field.) Turning the relation around, global optical outflow is said to be information about forward rectilinear motion of a point of observation in the sense that, given that there are no other natural ways of producing global optical outflow (Turvey, 1979), global outflow is specific to forward rectilinear motion. This is Gibson's way of defining the information contained in the light—it is optical structure lawfully generated by the layout of surfaces and by movements of the point of observation relative to the layout. We can capture the essence of the Gibsonian view in terms of Barwise and Perry's situation-type:

$$\text{Situation-type F} \quad \underset{\text{specifies}}{\overset{\text{lawfully generates}}{\rightleftharpoons}} \quad \text{Situation-type O.} \qquad (4)$$

Put very simply, under Gibson's ecological analysis, O means F.

In many circles, however, there is a reluctance to use the term 'means' or to construct a phrase of the form 'O's meaning is F' in the absence of a living thing, an agent. Thus, for Barwise and Perry a relation such as (4) is only one half of their theory of meaning as it might apply to a given animal. The other half is the attunement of the animal in question to the relations. For them, information and meaning are equated but the equation holds, strictly speaking, only when there is attunement of the animal. In short, in Barwise and Perry's situation semantics, the meaning of an event o that is of the event-type O is a product of the relation F ----> O and attunement to the relation. Putting a living thing that sees and locomotes (and which, therefore, must be attuned by definition) at the point of observation relative to an artificially generated outflowing global optical field, underscores the identity of information and meaning to which the ecological approach and situation semantics subscribe. If O means F, then for a human observer maintaining an upright stance, global optical outflow will induce backward postural adjustments since forward movement rather than vertical stasis is 'occurring.' Experimentally, this is shown to be so (Lishman & Lee, 1973; Lee & Aronson, 1974).

We are not fully comfortable with the notion that the relation of O and F can be talked about in two ways, (1) as 'O informs about F' in the absence of an attuned agent, and (2) as 'O means F' given an attuned agent. Our discomfort arises from the desire to develop a consistent direct realist position in perceptual theory (Michaels & Carello, 1981; Turvey et al., 1981; Shaw, Turvey, & Mace, 1982) and our recognition of how elusive this realist goal has been in the past. It may be a quibble but it seems to us that a realist perspective is undercut to the degree that we cannot talk clearly and confidently about situations and events as having meanings for the activities of organisms, regardless of the psychological states of organisms. Given the low-level development of the concept of attunement in situation semantics, there is a danger that attunement might be read in a psychologically contributory sense, viz., the organism is able to interpret the information, that is, to ascribe meaning to the information.

From a realist viewpoint, meanings are discovered by animals, not invented or created by them. The nomic structural constraints of ecological optics relate kinetic and kinematic facts at the ecological scale to optical structure. They are the sine qua non for the evolution of visually guided locomotion, whether the forces for locomotion be produced by legs, fins, wings, or machine (Gibson, 1979). To say that the lawfully produced optical properties are merely information fails to convey the existential import of the nomic constraints of ecological optics: They have been the basis for the successful locomotion of an indefinitely large number of species for a very long period of time. Indeed, we would speculate (and, we hope, not glibly) that attunement to these constraints could not have come about unless they were already meaningful, that is, unless the kinetic consequences of a (naturally occurring) global optical pattern always held.

To fix this equation of information (in the specificational sense) and meaning, consider a point of observation moving on a rectilinear path that is interrupted by a substantial surface perpendicular to the ground. The structured light to the point of observation is usefully construed as nested visual solid angles with the point of observation as their common vertex (Gibson, 1979). Crudely speaking, the larger solid angles correspond to the faces of surface layout and the smaller solid angles correspond to the facets. As a moving point of observation approaches the substantial surface on its path, the corresponding visual solid angles will dilate. Analysis shows that the inverse of the rate of dilation is a global property that is specific to the time-to-contact between the point of observation and the surface (Lee, 1976, 1980). To be somewhat pedantic, when a point of observation approaches a surface under constant force conditions (the kinetic perspective) and, therefore, at a constant velocity (the kinematic perspective), it defines a physical situation such that, for any distance between the point and the surface, there is a corresponding time before point and surface contact. The light is lawfully structured by this physical situation of imminent collision such that there are optical properties unique and specific to the facts that a collision will occur and that it will occur at a certain delay. We can identify the time-to-contact optical property, $\tau(t)$, then, in the terms of Barwise and Perry:

$$\text{Situation-type } \dot{\iota}(t) \quad \underset{\text{specifies}}{\overset{\text{lawfully generates}}{\rightleftharpoons}} \quad \text{Situation-type } \tau(t), \qquad (5)$$

where $\tau(t)$ means one thing and one thing only, namely, contact C, at so many seconds from now if the current conditions of motion persist. Contact will occur whether the point of observation is filled by an attuned agent, a blind agent, or a trolley. A given value of $\tau(t)$ means collision at a certain time. This fact of nature is the sort of meaningful invariant to which perceptual systems could adapt and become sensitive or attuned. That $\tau(t)$ is meaningful in the way we have suggested is shown in its use by gannets in controlling their diving for fish (Lee & Reddish, 1981), by flies in initiating their deceleration prior to contacting a surface (Wagner, 1982), and by humans in leaping to hit a ball (Lee, Young, Reddish, Lough, & Clayton, 1983).

As we have noted, $\tau(t)$ is information about an upcoming collision *if the current conditions of motion persist*. Obviously, the collision need not be inevitable if the conditions of motion are changed in appropriate ways, for example, if the point of observation stops or veers to the side. Moreover, the strength and timing of the collision can be controlled (as demonstrated by the examples above) if the point of observation accelerates or decelerates appropriately. Is there information for what is appropriate? For example, is there information specific to the circumstance 'deceleration is sufficient to come to a halt before contacting the surface'? Such control information is available in the first derivative of the time-to-contact variable, $d\tau(t)/dt$. In particular, if $d\tau(t)/dt \geq -0.5$, then deceleration is sufficient and there will not be contact; if $d\tau(t)/dt < -0.5$, there will be contact.

This 'type-of-contact' variable is of particular interest because it is a dimensionless quantity (i.e., it is not attached to any units of measurement) that distinguishes natural categories: contacts vs. noncontacts. The category boundary does not change--the meaning of the situation does not change--with changes in speed of the observation point, its distance from the surface, or the size of the surface. The information specifying the category boundary is lawfuly produced by the movement of a point of observation with respect to a surface. We have suggested elsewhere (Kugler, Turvey, Carello, & Shaw, in press; Turvey & Kugler, in press-a) that dimensionless quantities that mark off distinct specificational states play the same significant role in law-based explanations of the control of activity as dimensionless quantities that mark off distinct physical states play in law-based explanations of cooperative phenomena.

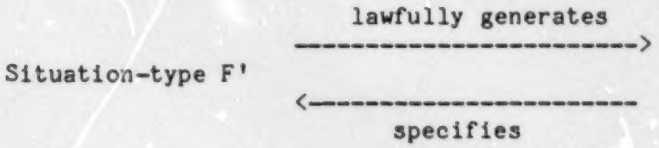
How Situation-type Meanings Become Situation Meanings

The above are examples of optical properties lawfully linked to particular relationships between a moving *point of observation* and a layout of surfaces. They are examples of nomic structural constraints that underwrite situation-type (or event-type) meanings for locomoting agents, if agents happen to be about. Building the laws of ecological optics around an unoccupied point of observation is an important move: The laws are thereby seen to be general and public, in that any observer, in principle, can occupy any point of observation and share with other observers over time the invariants in the ambient optic array to that point. Given the fact that these situation-type meanings are observer indifferent, however, we need not expect them to determine activity fully when an observer is brought into the picture (just as we do not expect the laws of motion by themselves--operating, as they must, within certain boundary conditions--to rationalize fully a given particle's trajectory). An occupant at a point of observation transforms a situation-type meaning into a situation meaning and while the latter depends on the

former, it is not identical with it—as Barwise and Perry take great pains to note.

We wish to show, as do Barwise and Perry, that there is nothing spooky about this transformation of situation-type meaning into situation meaning. An observer occupying a point of observation will have a magnitude (height, weight) that defines an intrinsic scale for the laws of ecological optics, and a repertoire of effectivities (goal-directed activities) that define the uses to which the information based on these laws is to be put. As it is with Barwise and Perry's discourse situations, connections, and resource situations, which squeeze different situation meanings out of an invariant linguistic situation-type meaning (underwritten by conventional structural constraints), so it is with scale and intention, which squeeze different situation meanings out of an invariant action situation-type meaning (underwritten by nomic structural constraints). We expect that, formally speaking, these two sets of 'boundary conditions' may have much in common. Let us concentrate, however, on examples of how scale and intention produce situation meanings.

The motion (F') of a point of observation over one surface towards a drop-off to another lower surface will lawfully generate an optical flow (O') distinguished by a discontinuity, viz., a horizontal margin above which optical structure magnifies and gains and below which optical structure magnifies but does not gain. This nomic structural constraint and the information in the specificational sense that it yields can be represented as:



(6)

The situation-type meaning of O' is 'approaching a brink'. If the point of observation is occupied, say, by a running, four-legged animal, then the situation-type meaning is too general and insufficiently constrains the animal's behavior. The richer, particular meanings of 'approaching a step-down place' or 'approaching a jump-down place' or 'approaching a falling off place' are required for the successful control of locomotion. These meanings are situation meanings. They depend on the magnitude of the brink relative to the size of the animal. What is a step-down place for one animal (e.g., a horse) is a jump-down place or a falling-off place for another animal (e.g., a mouse).

The orthodox move is to treat these situation meanings as subjective—that is, as mental categories imposed on an objective, meaningless surround. This is where the spookiness creeps in. Gibson's ecological realism and Barwise and Perry's situation semantics reject this move to subjective categories. Rather, the situation meanings in question must be underwritten by scaled nomic structural constraints; they are real relations between real properties of the animal-environment system to which the animal can become attuned.

The strategy, roughly speaking, is to note that (a) the magnitudes of surface layout are describable in units of the animal such as eye height or stride length; (b) above some critical number of a body-scaled unit such as n (eye heights), a drop-off cannot be negotiated by stepping down; (c) the optical flow can be shown to specify surface layout in body-scaled units (Lee, 1980); and (d) given (c), there is a dimensionless optical property like dτ(t)/dt that marks off at a critical value distinct specificational states, viz., 'approaching a step-downable place' and 'approaching a non-step-downable place' (Turvey & Kugler, 1984). That is, given the optical structure O' fashioned by any point of observation approaching any brink in a surface, there is a scale transform s effected by a particular animal a at the point of observation such that s(O')---->O", where O" is the optical structure specific to the brink in the scale of a. In Barwise and Perry's terms, O' is efficient--although its meaning is fixed ('brink'), its "interpretation" ('step-downable,' 'not step-downable') varies with s. To reiterate another central theme of situation semantics, "...efficiency is crucial to all meaning."

A similar scenario can be written for the role of intention in transforming situation-type meaning into situation meaning. For example, a baseball fielder bent on catching a fly ball transforms situation-type meaning 'impending collision' into 'thing to be intercepted,' while a boxer with a glass jaw transforms 'impending collision' into 'thing to be avoided.' Each intention defines a natural category (selects values of the final conditions of a law) such as 'hard contact,' which, in turn, specifies the activities that will produce that category (i.e., constrains values of the initial conditions of the law). Just as understanding the interpretation of an utterance requires understanding its context of use in Barwise and Perry's terms, so understanding how an intention transforms a situation-type meaning into a situation meaning requires understanding the context of laws under which the intention brings the animal, including the convention that defines the initial conditions to be assumed given the final conditions to be obtained (see Turvey et al., 1981, for a more thorough discussion of this line of reasoning).

Throughout Gibson's ecological realism and Barwise and Perry's situation, semantics is a commitment to treat meaning as an aspect of reality. This sort of treatment gives rise to explanations of meaning that appeal to natural law; understanding meaning is not qualitatively different from understanding other natural phenomena. The strategy is reinforced, for the Gibsonian program, in never losing sight of the control of locomotion as the paradigmatic problem to be understood. Locomotion is a skill that is not limited to humans and, therefore, the temptation to ascribe it to special mental powers is lessened. Barwise and Perry, on the other hand, are trying to be realists in a bailiwick where mentalese is at its most alluring. We applaud their efforts.

References

Dretske, F. (1981). Knowledge and the flow of information. Boston: Bradford Books.

Gibson, J. J. (1954). A theory of pictorial perception. Audio-Visual Communications Review, 1, 3-23.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). The ecological approach to visual perceptions. Boston: Houghton-Mifflin.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. E. (in press). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren, Jr. & R. E. Shaw (Eds.), Persistence and change: Proceedings of the first international conference on event perception. Hillsdale, NJ: Erlbaum.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Lee, D. N., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. Perception & Psychophysics, 15, 529-532.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. Nature, 293, 293-294.

Lee, D. N., Young, D. S., Reddish, P. E., Lough, S., & Clayton, T. M. H. (1983). Visual timing in hitting an accelerating ball. Quarterly Journal of Experimental Psychology, 35A, 333-346.

Lishman, R., & Lee, D. N. (1973). The autonomy of visual kinaesthetisis. Perception, 2, 287-294.

Michaels, C. F., & Carello, C. (1981). Direct perception. New York: Prentice-Hall.

Reed, E. S. (1981). Indirect action. Unpublished manuscript, University of Minnesota, Center for Research in Human Learning.

Shannon, C. E. (1949). The mathematical theory of information. Urbana, IL: University of Illinois Press.

Shaw, R. E., Turvey, M. T., & Mace, W. (1982). Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes (II). Hillsdale, NJ: Erlbaum.

Turvey, M. T. (1979). The thesis of efference-mediation of vision cannot be rationalized. The Behavioral and Brain Sciences, 2, 81-83.

Turvey, M. T., & Kugler, P. N. (1984). An ecological approach to perception and action. In H. T. A. Whiting (Ed.), Human motor actions---Bernstein reassessed. Amsterdam: North Holland.

Turvey, M. T., & Kugler, P. N. (in press). A note on equating information with symbol strings. American Journal of Physiology.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). Cognition, 9, 237-304.

Wagner, H. (1982). Flow-field variables trigger landing in flies. Nature, 297, 147-148.

Footnote

[1]Although we agree with the distinction that Barwise and Perry draw between situation-type and event-type, for purposes of exposition we use situation-type for both circumstances.

A COMMENT ON THE EQUATING OF INFORMATION WITH SYMBOL STRINGS*

M. T. Turvey+ and Peter N. Kugler++

Physicists have been included in the past to regard "information" as a physical variable similar in kind to energy or matter (e.g., Layzer, 1975; Tribus & McIrvine, 1971). There are objections, however, to carrying this inclination over to the realms of biology, physiology, and psychology. The equating of "information" with negative entropy or an absolute measure of objective order does not adequately capture the ways in which the term "information" is used in explanations of the phenomena characteristic of those realms. There is a very general impression that the various explanatory roles ascribed to "information" in biology, physiology, and psychology are performable by symbols organized by a grammar.

What are the roles that symbol strings fulfill? Fundamentally, they are indicational and injunctional: Symbol strings can indicate states of affairs (e.g., "deficiency of metabolite so-and-so"; "road work ahead") and they can direct or command states of affairs (e.g., "release hormone so-and-so!"; "slow down!"). This popular quasi-linguistic view of "information"—what might be termed the indicational/injunctional sense of information (cf. Reed, 1981; Turvey & Kugler, 1984)—is central to the papers of Bellman and Goldstein, and Iberall. Our efforts in this brief note are directed at putting this indicational/injunctional sense of information into perspective. Insofar as the issue of the continuity of linguistic and movement capabilities involves the concept of information, clarifying the different senses of the concept, and their relationship, will prove helpful. Two rather different sets of arguments are involved—those attributed to Howard Pattee and those attributed to James Gibson.

Pattee (1973, 1977) has identified two modes of complex system functioning: A discrete mode characterized as rate-independent operations on a finite set of symbols, and a continuous mode that refers to the rate-dependent interplay of dynamical processes. Given this distinction, one can ask how symbol strings and dynamics coevolve from the cellular level up through the evolutionary scale. More pointedly, the question can be raised: Are there universals of symbol string/dynamics interactions that might be appropriate to an understanding of the linguistic and coordinated movement capabilities of living systems? Pattee addresses these questions through the problem of

---

*American Journal of Physiology, in press.
+Also Department of Psychology, University of Connecticut.
++Also Crump Institute for Medical Engineering, University of California, Los Angeles.
Acknowledgment. The writing of this paper was supported in part by ONR Contract N0014-83-C-0083 awarded to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

enzyme folding. This particular example consists of two qualitatively different phases: the genetic code synthesizes an amino acid string, which then folds into a functioning enzyme. The translation of the DNA symbols into animo acid strings is a discrete symbolic process, while the folding of the one-dimensional amino acid string into a three-dimensional machine is a continuous dynamical process. The former is a constraint on the latter. To describe the relationship as one of constraint is an important step for Pattee, for it suggests that the system's meaning--its dynamic ability--does not merely reduce to a symbolic representation. The symbolic mode harnesses the forces responsible for the function, but the symbolic mode is not equated with the function. But neither is the dynamic mode completely autonomous. The folding of the enzyme cannot proceed until the code provides the necessary constraint. In other words, neither mode alone is sufficient for the activity in question.

Of significance is the observation that the discrete symbolic mode, information in the indicational/injunctional sense, is kept to a minimum in natural systems (Pattee, 1980). Information construed quasi-linguistically does not provide all of the details for a given process; it acts as a constraint, of the nonholonomic type, on natural law so that the dynamic details take care of themselves. In other words, by Pattee's analysis, most of the complex behavior of living systems is essentially self-assembly that is "set up" by symbol strings, but not explicitly controlled by them. Presumalby this should be no less true of the linguistic and movement coordination capabilities of biological systems. For Pattee, complete comphrehension cannot be gained by appealing to symbol-string processing or to physics alone. Both must be used together, but in a special way. Pattee advises: Use physics cleverly so that symbol strings need only be used sparingly in order to assure the parsimony of the explanation.

As noted, symbol strings are incomplete--they are limited in detail with respect to the detail of the processes that they indicate or direct. A number of perplexities are generated by this incompleteness. For example, on what grounds and by what means does a particular symbol string get created rather than another, referring elliptically to one set of properties of the indicated or directed dynamical process rather than another? What determines the detail of the indicated or directed dynamical process that the symbol string represents? Taken together, these two questions require an answer beyond that given by a physics (e.g., Prigogine's Dissipative Structure Theory, Iberall's Homeokinetics) that seeks to explain how structure evolves with a consequent loss of dynamical degrees of freedom. What is required is an explanation of how that loss is <u>special</u> yielding a symbol string, an alternative description (Pattee, 1972), that is <u>privileged</u> with respect to the dynamical process that it indicates or directs (see Carello, Turvey, Kugler, & Shaw, 1984). There are shades of the problem of induction (Goodman, 1965) here, the problem of projectable predicates or properties, which continues to resist solution in conventional philosophy and psychology. Consider another consequence of incompleteness. Because of its necessarily reduced detail, a symbol string cannot specify a process or act, that is, it cannot provide a lawful basis for the process. This is not to say that information in the indicational/injunctional sense cannot be responsible for a process in part, only that it cannot constrain a process in full. Pattee's paradigmatic example is meant to suggest that the known laws of physics complete the picture--filling in what the symbol string leaves out. But we doubt whether all relevant examples succumb to this solution, <u>tout court</u>. It seems to us that in many (if not most) bio-

logical settings the dynamical details "take care of themselves" because there is non-symbolic information that specifies how they should do so. As Iberall and Soodak (in press) express it, a cooperativity is a state of affairs of an ensemble that is maintained from below by the activity of the atomisms of the ensemble and from above by the field boundary conditions (equated with nonholonomic constraints qua symbol strings in most biological instances). The intimation below is that cooperativities involving biological atomisms are predicated in large part on information in a non-symbolic sense that is made available in the course of atomistic activity.

Gibson's (1966, 1979; Reed & Jones, 1982) focus has been the control of locomotory activity in natural cluttered surroundings. His definition of information is explicit and distinct from the orthodox sense of information as indicational/injunctional. For Gibson, information in the case of vision is optical structure that is lawfully generated by environmental structure (the layout of surfaces) and by movements of the animal (both movements of the limbs relative to the body and movements of the body relative to the surround). The optical structure does not resemble the facts of the animal-environment system, but it is specific to them, in the sense of being lawfully dependent on them. In short, Gibson's sense of information is specificational. A simple example illustrates the relation between the two senses of information, Gibson's and the orthodox. Symbol strings on the highway of the type SLOW DOWN and STOP are intended to direct the dynamics of traffic flow. For atomisms (humans) that can read the symbol strings, complying with these injunctions is possible only if there is continuously available information specific to the retardation of forward motion and the time to contact with the place where velocity is to go to zero. A deceleration of global optical outflow specifies the slowing down of a moving point of observation relative to the persistent, non-moving layout of surrounding surfaces. The inverse of the rate of dilation, of the visual solid angle to the point of observation that is created by approach to the place where motion is to be fully arrested, specifies continuously the time at which the place will be contacted. And the first derivative of the time-to-contact optical property specifies that the forward motion will or will not be arrested in time under the current conditions (forces) of motion. (See Gibson, 1979; Kugler, Turvey, Carello, & Shaw, in press; Lee, 1980, for a detailed discussion of each of these forms of specification.) This example suggests that without information in the specificational sense, information in the indicational/injunctional sense is impotent. Further, this example suggests that for a given process, the degree of detail in a symbol string is inversely related to the availability of information in the specificational sense. At the very least, the information available in the specificational sense determines the lower bound on the detail of information in the indicational/injunctional sense.

Stated in more general terms, Gibsonian information is a physical variable that can be identified with low-dimensional macroscopic properties of low-energy fields lawfully generated by properties of system-and-surround (Kugler et al., in press). For a system that has an on-board source of available potential energy (such that it can resist the surround's forces through the generation of forces of its own), information in the Gibsonian specificational sense is the basis of the system's coupling to its surround. Where a convention, abstractly interpreted, leads the system to take a nongeodesic path (route), information in the specificational sense provides the support by which this elected activity is made possible.

In summary, the points we wish to underscore are these: (1) the indicational/injunctional sense of information is not exclusive; (2) information in the indicational/injunctional sense is predicated on information in the specificational sense; and (3) the perplexities surrounding the incompleteness of symbol strings may be dismissed in a principled fashion by a thoroughgoing analysis of information in the specificational sense (cf. Carello et al., 1984; Turvey & Kugler, 1984).

References

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1984). Inadequacies of the computer metaphor. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum.

Gibson, J. J. (1960). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Goodman, N. (1965). Fact, fiction and forecast. Indianapolis, IN: Bobs-Merrill.

Iberall, A. S., & Soodak, H. (in press). A physics for complex systems. In F. E. Yates (Ed.), Self-organizing systems: The emergence of order. New York: Plenum.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. E. (in press). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren & R. E. Shaw (Eds.), Persistence and change: Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum.

Layzer, D. (1975). The arrow of time. Scientific American, 223, 56-59.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Pattee, H. H. (1972). Laws and constraints, symbols and language. In C. H. Waddington (Ed.), Towards a theoretical biology. Chicago: Aldine.

Pattee, H. H. (1973). Physical problems of the origin of natural controls. In A. Locker (Ed.), Biogenesis, evolution, homeostasis (pp. 41-49). Heidelberg: Springer-Verlag.

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. International Journal of General Systems, 3, 259-266.

Pattee, H. H. (1980). Clues from molecular symbol systems. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie.

Reed, E. (1981, November). Indirect action. Unpublished manuscript, Center for the Study of Human Learning, University of Minnesota.

Reed, E., & Jones, R. (1982). Reasons for realism: Selected essays of James J. Gibson. Hillsdale, NJ: Erlbaum.

Tribus, M., & McIrvine, E. C. (1971). Energy and information. Scientific American, 224, 179-186.

Turvey, M. T., & Kugler, P. N. (1984). An ecological approach to perception and action. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. Amsterdam: North Holland.

AN ECOLOGICAL APPROACH TO PERCEPTION AND ACTION*

M. T. Turvey+ and Peter N. Kugler++

1.0 Introduction

In his chapter on "Some emergent problems of the regulation of motor acts" Bernstein (1967) identifies four major problems:

(1) If perceiving were not a matter of being accurately aware of the objective facts of the environment and of one's actions, then the reliable control of activity would not be possible. However, the orthodox theory of receptor processes implies an arbitrary relation between these processes and the circumstances--environment and action--to which they nominally refer. This theory is inadequate to explain the everyday achievements of animal activity. What is needed is a theory that accounts for how perceiving keeps an animal in contact with the reality that bears on the successful conduct of its actions.

(2) Patently, animal activity is an instance of self-regulation, but what kind of self-regulation? Is it of the type conventionally expressed by self-regulating artifacts or do the regularities of animal activity follow from principles that are, as yet, unique to natural systems?

(3) Neither the geometry nor the kinematics of movement can serve, in the general case, as the determinant of the composition of an act. An action is what it is by virtue of its intention, that is, the motor problem (a needed change in the relation of the animal and its environment) toward which the action is directed as a solution. How are we to understand an intention as (a) the principle guiding the overall formation of an act and (b) the influence dominating the selection of its details?

(4) Clearly the control of activity is more than a retrospective matter. In the most general of cases, control must be prospective. For example, in basketball, one exerts forces against the ground of a specific magnitude so as to cause the hands to be at a specific height at a specific time to intercept a thrown ball. What is the basis of this anticipatory capability that makes possible the realization of any goal-directed activity?

---

*Also in H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. Amsterdam: North-Holland Publishing Company, 1984.

+Also University of Connecticut.

++The Crump Institute for Medical Engineering, University of California, Los Angeles.

Acknowledgment. The preparation of this paper was supported in part by ONR Contract N00014-83-C-0083 awarded to the Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

Problems (1) and (4) are discussed in Section 2.0 and problems (2) and (3) are discussed in Section 3.0.

2.0 On the Objectivity and Accuracy of Perceiving

For any animal, activity takes place with respect to surfaces. For terrestrial animals, the most important surface is the ground. The ground is not even. Neither is it geometrically and materially uniform from place to place. There are gradual and sharp changes in the ground level. There are cracks and gaps. Liquid and solid areas are interspersed. Further, the ground surface is cluttered with closed, substantial surfaces. Some of these are attached, others are movable and some move under their own power. The clutter varies greatly in size. But for any terrestrial animal there are always closed, substantial surfaces both smaller and larger than its size. Some of the ground's clutter are barriers to locomotion, but invariably there are gaps large enough to permit passage and barriers small enough to be hurdled or climbed. Locomoting from place to place, finding paths through the clutter, is necessary given the uneven distribution of the resources on which the persistence of the animal depends.

As Bernstein remarks, the meaningful problems that activity solves arise out of the layout of surfaces surrounding the animal, the environment. A few such meaningful problems are depicted in Figure 1. Awareness of the "problems" and awareness of the activities that do or do not solve them is the role of perceiving. It is obvious to Bernstein that perceiving (both the layout of surfaces and activities with respect to the layout) must be "objective" and "accurate." If perceiving fell short of these requirements--if it were, on the contrary, "subjective" and "inaccurate"--then meaningful, adaptive activity would not be possible. Bernstein writes (1967, p. 117): "We may consider the formulation of the motor problem, and the perception of the object in the external world with which it is concerned as having their necessary prerequisites in maximally full and objective perception both of the object and of each successive phase and detail of the corresponding movement which is directed towards the solution of the particular problem." What Bernstein says seems straightforward enough: perceiving must keep an animal in contact with its surroundings and with its behavior. It will be argued, however, that a number of fairly radical steps have to be taken to insure that the theory of perception that we develop as scientists can live up to the natural demands placed on perception by normal activity in cluttered surroundings.

Clearly, Bernstein believes that the role of "afferentation" in the guidance of activity is the significant role, even though afferentation as a trigger of reflexes has a better scientific pedigree and is better understood by physiologists. The triggering role of afference assumed prominence because of the tendency to focus research on artificial movements--discrete responses made to momentary and punctate stimuli--rather than on activities resolving environmentally defined problems. Bernstein considered this triggering role of afference, developed as it was in the context of the reflex arc, to be overvalued and pointed out two unwelcome consequences of this overvaluation.

First, it established a bias to equate receptor processes in general with signals that release or inhibit reactions. Bernstein reminds us that this equation leads to the unacceptable interpretation of the receptor processes accompanying linguistic events as just triggers--the so-called second-signaling system. Closer to the present concerns, he points out that the emphasis

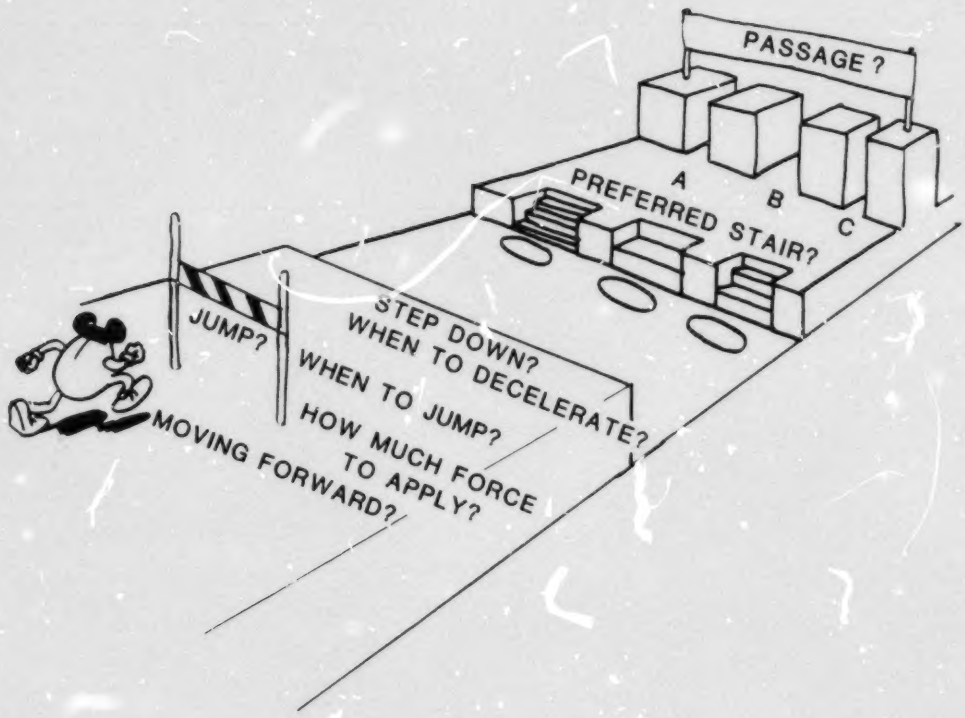


Figure 1. A small sample of the meaningful problems that the surrounding layout of surfaces poses for a locomoting animal.

on the afferent triggering of reactions obscured the fact that afference modulates ongoing movements. Bernstein saw a distinction between the traditional physiology of reaction and the physiology he wished to promote--a physiology of activity (Gelfand, Gurfinkel, Fomin, & Tsetlin, 1971; Reed, 1982a). In this respect (and others) he was of kindred spirit with Gibson (1966, 1979, 1982) in rejecting the classical view of action as (merely) responses triggered by signals emanating from either outside the body or inside the brain. However, it should be noted that Bernstein labored under a major terminological manifestation of the classical view, namely, the correspondencies of the terms "sensory" and "afferent," "motor" and "efferent." This was unfortuante given that he rejected the conceptual identity that these correspondences implied.

As Gibson (1966, 1979, 1982) and Reed (1982a) have ably argued, the psychological concepts of sensory and motor cannot be equated, respectively, with the anatomical structures termed afferent and efferent. The anatomical definition of the sensory system (as receptor elements, cortex, and the afferent pathways that mediate them) fails to accommodate the adjusting, optimizing, steering, and symmetricalizing of sense organs—that is, their purposive activity (Gibson, 1966). Bernstein recognized this inadequacy. In referring to the systematic searching by sense organs, he wrote (p. 117): "This is an entirely active process, and the effector side of the organism is here employed in a manner completely analogous to that which is later explained to underlie afferentation in the control of movements." The anatomical definition of motor system (as cortex, motoneurons, and the efferent pathways that

mediate them) fails to accommodate the dynamic responsiveness of effector organs to changes in the external force field brought about by changes in the orientation of effectors to the surround and to the body--that is, their contextual sensitivity (Turvey, Shaw, & Mace, 1978). More than anybody before him, Bernstein sought to substitute the analysis of action in terms of efferent commands from cortex to motoneurons by an analysis of action as the selective use of information about the environment and about one's movements to selectively modulate one's movements with respect to the environment (cf. Gibson, 1979).

Second, the tradition of regarding afferents as triggering signals enforced, in Bernstein's view, a general attitude toward afference as arbitrarily related to the environmental conditions that cause it. All that is required for the successful initiation of a reflex is afferentation that is constant and recognizable by the effector apparatus. The proximal cause of a reaction need bear no necessary relation to the distal cause. As Bernstein sees it, this idea of an arbitrary connection between afferent states of affairs and environmental states of affairs is pernicious. If the afferent (or sensory) codes are arbitrary (as is claimed by Müller's Doctrine of Specific Nerve Energies and its successors) and if what the animal perceives is based on these codes, then what is there to guarantee that the animal's perception is objective and accurate? The depth of Bernstein's concern is expressed in this quotation (p. 126): "...from the fact that it is clearly possible to reconcile the perfect operation of reflex functions with the complete arbitrariness of their sensory codes it is very easy to slide from the position of the recognition of the symbolic nature of all reception in general, and of the conditionality of the picture of the world in the brain and the psyche, to the concept of unknowability of objective reality and similar idealistic conceptions..."

2.1 The Cartesian Program

The orthodox and very popular representational/computational approach to mind (see Chomsky, 1980; Fodor, 1975; Pylyshyn, 1980) is consistent with the arbitrary coding theme that Bernstein believes (incorrectly, as we will claim below) to be rooted in the reflex philosophy and methodology. The representational/computational view abides by a "formality condition"--the explicit understanding that mental operations are formal, symbol manipulations performed on formal, symbol structures (Fodor, 1980). To a computer (and, by analogy, to a brain) it is immaterial whether its internal codes refer to this or that fact; how the signals are formatted and how they relate consistently among themselves by rule are what matters, not their meaningful content. We raise the spectre of the formality condition for two reasons. One reason is that Bernstein, despite his dislike of this condition in the guise that was familar to him, invokes a mechanism for the control of activity that is continuous with the representational/computational thesis and, therefore, with the formality condition. Bernstein suggests that an ordered sequence of set points--representations of required values--governs the flow of afference and efference within the acting animal. The ordered sequence is a program prescribing the general form of the activity; it is a representation of the activity for the effector organs (cf. Cummins, 1977; Shaw, Turvey, & Mace, 1982).

The other reason is that the formality condition is clearly tied to the historical tradition that began with the Cartesian Doctrine of Corporeal Ideas. It is this tradition that encourages the arbitrary coding interpretation of afference, not the reflex arc methodology, which is itself a restatement of the Cartesian doctrine (Reed, 1982b). Descartes' doctrine, stated very generally, is that all awarenesses are awarenesses of states of the body or, as we would be more prone to say today, states of the brain. In contemporary thought, it is said that direct access to environmental and behavioral states of affairs is limited to the physical (or bodily) outputs of transducers that are linked not to the environmental and behavioral states, but to the basic energy variables, e.g., intensity and wavelengths of light (Boynton, 1975; Fodor & Pylyshyn, 1981). The question that this doctrine poses has been at the base of almost all theories in psychology, viz.: How can the environment be known objectively and accurately and acted upon successfully when the ideas one has about such things are based on awarenesses merely of brain states? Descartes had an answer to the subsidiary question of how primary objective qualities might be derived from secondary subjective qualities and it has been a persistent ingredient in almost all subsequent theorizing. He assumed an act of understanding that passed judgment on what environmental things might have caused the brain state; in his best known example, he assumed a rule-governed, quasi-mathematical process of inference from the states of the eye muscles and the visual nerves to the distance of an environmental object.

We can now focus sharply on the full implications of Bernstein's innocent claim that the coordination of an animal and its environment must be based on objective and accurate facts. Because of the pervasiveness of the Cartesian doctrine in physiology, psychology and cognitive science (see Reed, 1982b; Shaw et al., 1983), it is generally accepted that an animal's awareness of its activities and of the surface layout to which they refer is not direct but mediated. Descartes had proposed rules, inferences and judgments to get to these objective facts of activity and environment from the directly given, subjective brain states. To Descartes' list of cognitive or epistemic mediators, later theorists have added representations, schemas, programs, models, organizing principles, meanings, concepts, and the like. Whether dressed in its traditional or modern garb, the Cartesian program for explaining how felicitous activity is achieved in a cluttered environment faces a profound predicament. There is nothing in this explanation to guarantee that the proposed inferential operations performed on the brain states will yield conclusions that are objective and accurate rather than fatuous. In responding to John Locke's version of the Cartesian program, Berkeley thought a guarantee was unwarranted and emphasized the phenomenalism (that there are only phenomenal objects such as ideas) implicit in the Cartesian program. Hume thought a guarantee was unlikely to be forthcoming and emphasized the skepticism (that there may well be an environment and activities oriented to it, but no one can be sure of their existence) implicit in the Cartesian program. It is to thoughts such as those expressed by Berkeley and Hume that Bernstein refers when he remarks on "...the concept of unknowability of objective reality and similar idealistic conceptions..." Bernstein (p. 125) goes on to say (too cavalierly, in our opinion) that such thoughts "...have been disproved by authentic science long ago." As scientists committed to an objective reality, we must claim that it is knowable by animals, more or less. However, a scientific account of perception that is consistent with this realist posture has been thwarted, in our view, by the almost universal acceptance of the Cartesian program. As long as the Cartesian program is the accepted strategy

for explaining the coordination of an animal and its environment--as long as the awareness of surface layout and action is claimed to be cognitively mediated--then the thoughts of Berkeley and Hume cannot be dismissed cavalierly and the predicament identified above remains firmly entrenched in psychological and physiological theory.

Accurate, objective conclusions might be assured if the inferential operations (and the various cognitive entities such as representations, etc.) were tightly constrained by reality. But the Cartesian program denies an animal direct contact with reality; to reiterate, only brain states are directly contactable. The problem for the Cartesian program, therefore, is how to get the reality that bears on the felicitous control of activity into the mind or nervous system of the animal. There are several responses to this problem. The most popular response is that a model of reality is constructed by a process of justifying inferences in the course of either evolution or ontogeny, or both (Bernstein advances a solution of this type). We will briefly summarize some of the reasons that render this response (scientifically) unacceptable (see Shaw et al., 1982; Turvey, Shaw, Reed, & Mace, 1981, for a fuller discussion).

All forms of non-demonstrative inference proposed by inductive logicians--enumerative inference, eliminative inference, and abductive inference--can be expressed as a confirmatory relation between evidence and hypothesis. The conditions of adequacy for confirmation vary among the forms of inference (see Smokler, 1968), but this is immaterial to the points we wish to make, viz., that the very notion of inference requires (1) the ability to project relevant hypotheses and (2) the availability of predicates in which to frame evidence statements and hypotheses. To clarify, the notion of a basic set of hypotheses is explicit in eliminative and abductive inference and implied in enumerative inference. For example, one version of abduction (Hanson, 1958, p. 72) goes as follows:

Some surprising phenomenon P is observed.

P would be explicable as a matter of course if H were true.

Hence, there is reason to think that H is true.

If a model of reality were derived from inference, then it would have to be supposed that appropriate hypotheses--hypotheses that are generalizations about environmental states of affairs--were already at the disposal of the animal. What is their origin? Surely the answer cannot be "inference," for that would precipitate a vicious regress. But if the answer is not "inference," then the only option for the Cartesian program is that the origin of the hypotheses is both extra-physical and extra-conceptual. These are mutually exclusive categories.

The same conclusion follows from the point about the availability of two kinds of predicates, those for framing evidence statements and those for framing hypotheses. The predicates in an evidence statement (the outputs of the transducers) stand for energy variables and, by argument, have their origin in physical processes. But for any form of inference there must be available, concurrently, predicates in which to couch hypotheses and these must be predicates that stand for environmental properties (such as an obstacle to locomotion). The origin of these environment-referential predicates cannot be

inferential, otherwise the argument is regressive; and it cannot be physical (law-based) because that option is denied the Cartesian program, by definition.

The general conclusion to be drawn is that a reliance on inference takes out a loan of intelligence that science can never repay: The Cartesian program is not a scientifically tractable program, and a fortiori, is a program for perception that science would be ill-advised to pursue.

2.2 Gibson's Ecological Program

We believe that the Cartesian program must be abandoned if a scientifically acceptable account is to be provided of the perceptual objectivity that Bernstein regards as the sine qua non of action. To ease the break with tradition, it may help to remember that Descartes built his perceptual theory around thought, not action. Gibson's (1979) is an approach to perceiving that takes the control of activity as its central concern. In this approach the Cartesian doctrine of corporeal ideas is rejected together with the many perplexities that it entails. Rather than founding perceptual theory on brain states that are related tenuously to the environments and activities of animals, Gibson founds perceptual theory on structured energy distributions that are lawfully related to the environments and actions of animals. Rather than asking how accurate objective inferences from brain states to the facts of environments and actions are made, Gibson asks how information specific to the facts of environments and actions is detected. Rather than assuming that the conventional variables of physics provide the only legitimate basis for describing the environment, Gibson advances the idea that the environment can be legitimately described in terms that are referential of the activity capabilities of animals.

It will not be possible for us to do complete justice to Gibson's perceptual theory in these pages (see Gibson, 1979; Michaels & Carello, 1981; Reed & Jones, 1982; Turvey et al., 1981; Turvey & Carello, 1981). We will restrict ourselves, therefore, to those Gibsonian concepts that we take to be most central to the control of activity--the concepts of information and affordance. And we will restrict ourselves to the perceptual system of greatest relevance--the visual perceptual system.

Information is optical structure generated in a lawful way by environmental structure (for example, surface layout) and by the movements of the animal, both the movements of its body parts relative to its body and the movements of its body as a unit relative to the environment. This optical structure does not resemble the sources that generate it, but is specific to those sources in the sense that it is nomically (lawfully) dependent on them. The claim is that there are laws at the ecological scale that relate optical structure to properties of the environment and action (Gibson, 1979; Turvey et al., 1981).

This treatment of information and the notion of ecological laws rests on an optical analysis that departs from the classical geometric ray optics and the more contemporary physical optics. Though some have argued to the contrary (e.g., Boynton, 1975; Johansson, 1970), neither of these analyses is sufficient to capture the richness of light's structure subsequent to multiple reflections from surfaces of varying inclination and substance and undergoing various types of change. Gibson's push has been for a theory of optics that

can do justice to ambient light as a basis for the control of activity. Given that activity is at the ecological scale of animals and their environments, Gibson termed the sought-after optical theory ecological optics. The limitations of conventional optical analyses recognized by Gibson (1961, 1979) are echoed by illumination engineers (e.g., Gershun, 1939; Moon, 1961; Moon & Spencer, 1981) whose goals are much more modest than Gibson's. In the subsection that follows we consider the activity-relevant questions raised by Figure 1 in terms of Gibson's ecological optics.

2.2.1 How does the animal know that it is moving forward?

Forward rectilinear motion of a point of observation relative to the surroundings will lawfully generate an expanding optical flow pattern globally defined over the entire optic array to the point of observation. [A locally defined expansion pattern, kinematically discontinuous at its borders with the optical structure in the large, would be lawfully determined by a part of the surround moving relative to the point of observation. In natural circumstances there can be no ambiguity, contrary to the standard claim (von Holst & Mittelstädt, 1950), about what is moving--the animal or part of its environment (Turvey, 1979).] As noted above, the lawfulness of optical structure at the ecological scale is the basis for its functioning as information for the control of activity: If A lawfully generates B, then B specifies A. Lishman and Lee (1973) have shown that humans walking voluntarily forward will report that they are walking backwards when exposed to the global optical transformation that is lawfully generated by backward locomotion (and which, therefore, specifies that the walker is moving backward). Further, when flying insects are exposed to global optical transformations that are the lawful consequences of forces that produce rotation, vertical displacement, and yaw, they respond with the appropriate counteracting forces (Srinivasan, 1977; Turvey & Remez, 1979).

2.2.2 How does the animal know from where to jump (to accommodate an upcoming barrier) and How does it know whether its deceleration is adequate (to accommodate its locomotion to an upcoming brink in the ground)?

The answers to both of these questions depend, in the Gibsonian perspective, on information about the imminence of contact (with barrier or brink). Lee (1974, 1976, 1980) and others (e.g., Koenderink & van Doorn, 1981) have identified an optical variable, symbolized as $\tau(t)$ by Lee (1976), that is equal to the inverse of the rate of dilation of a bounded region of optical structure. $\tau(t)$ is lawfully generated by the approach at constant velocity of a point of observation to a substantial surface in the frontal plane, or vice versa; it specifies the time at which the point of observation will make contact with the surface.

Obviously, the existence of an optical variable specifying time-to-contact bears directly on the question posed by Bernstein (Question 4 above) of how control can be prospective. Any answer to the question of prospective control is constrained by the requirements that (1) causes precede effects and (2) causes be actual rather than possible states of affairs. An event at a later time cannot cause an action at an earlier time and only actual events can be causal. In the Cartesian program, the bases of prospective control are representations; actual mental states existing in the present (rather than future, possible states of the animal-environment system) are the causes of activity. The logical format of these representations in the case of con-

trolled collisions must be that of a counter-factual, roughly of the form "if I don't change what I am doing and the conditions continue to be as they are, then X is likely to occur." The basis of prospective control in the Gibsonian program is exemplified by the time-to-contact variable, viz., there is information in the present optical structure (e.g., the value of $\tau(t)$ at ti) specific to what will occur if the present conditions continue (e.g., collision at time ti). To draw the contrast sharply, in the Gibsonian program the basis for prospective control is sought in laws at the ecological scale (that relate present optical properties to upcoming properties of the animal-environment system); in the Cartesian program the basis is sought in inferential processes (that relate the semantically neutral outputs of transducers to a counterfactual representation). Reiterating the arguments raised above, the Cartesian solution to the problem of prospective control begs the interesting questions; for example, how does the animal construct just that counterfactual representation that is right for the current situation?

Let us look at an example of the use of the time-to-contact variable. The gannet, a large seabird that feeds on fish, hovers about thirty meters above the water. On sighting a prey, it dives down first with its wings partly spread for steering and then with its wings folded so that it enters the water vigorously but cleanly. It may hit the water at speed approaching 25 ms^{-1} (or 55 miles h^{-1}). The action problem for the gannet is to retract its wings soon enough to avoid fracturing them but not so soon as to hinder the accuracy of its dive. Given that the gannet dives from varying heights, at varying initial speeds, and in varying wind conditions, how does it properly control its entry? Lee (1980) and Lee and Reddish (1981) have concluded that wing retraction is initiated when the time-to-contact variable reaches a certain margin value. (Because the animal is constantly accelerating due to gravity in the dive, the same margin value of $\tau(t)$ will be associated with different actual times-to-contact. The birds are seen to fold their wings a longer time before contact the higher the starting point of the dive.)

There is reason to believe that the time-to-contact variable is the basis of prospective control in a number of related circumstances. Data on the kinematics of catching a ball (Sharp & Whiting, 1974, 1975), hitting a baseball (Hubbard & Seng, 1954), infants reaching for a moving object (von Hofsten & Lindhagen, 1979; von Hofsten, 1983), stepping down (Freedman, Wannstedt, & Herman, 1976), and falling on one's hands against an inclined board (Dietz & Noth, 1978) are more or less amenable to such an analysis (see Fitch, Fuller, & Turvey, 1982; Fitch & Turvey, 1978; Lee, 1980). The last situation is depicted in Figure 2. The triceps brachii muscles are shown to tense in preparation for an upcoming collision in which the arms must absorb the momentum. With the eyes closed, the electromyographic index of the initiation of muscle tension is tied to the start of falling; with the eyes open and with different falling distances the index occurs at varying times after the start of falling but at an approximately constant time prior to contact.

We should remark that the fact of a simple, single optical property specifying the imminence of contact has implications for another of Bernstein's concerns, namely, how an animal can adjust its behavior to the velocity of things. Bernstein pursues a conventional argument that velocity is arrived at by a process of comparing the present location of a thing with the memory trace of an immediately preceding location and dividing the deduced distance traveled by an internally determined estimate of elapsed time. The inadequacies of this kind of explanation have been discussed in detail (Gib-

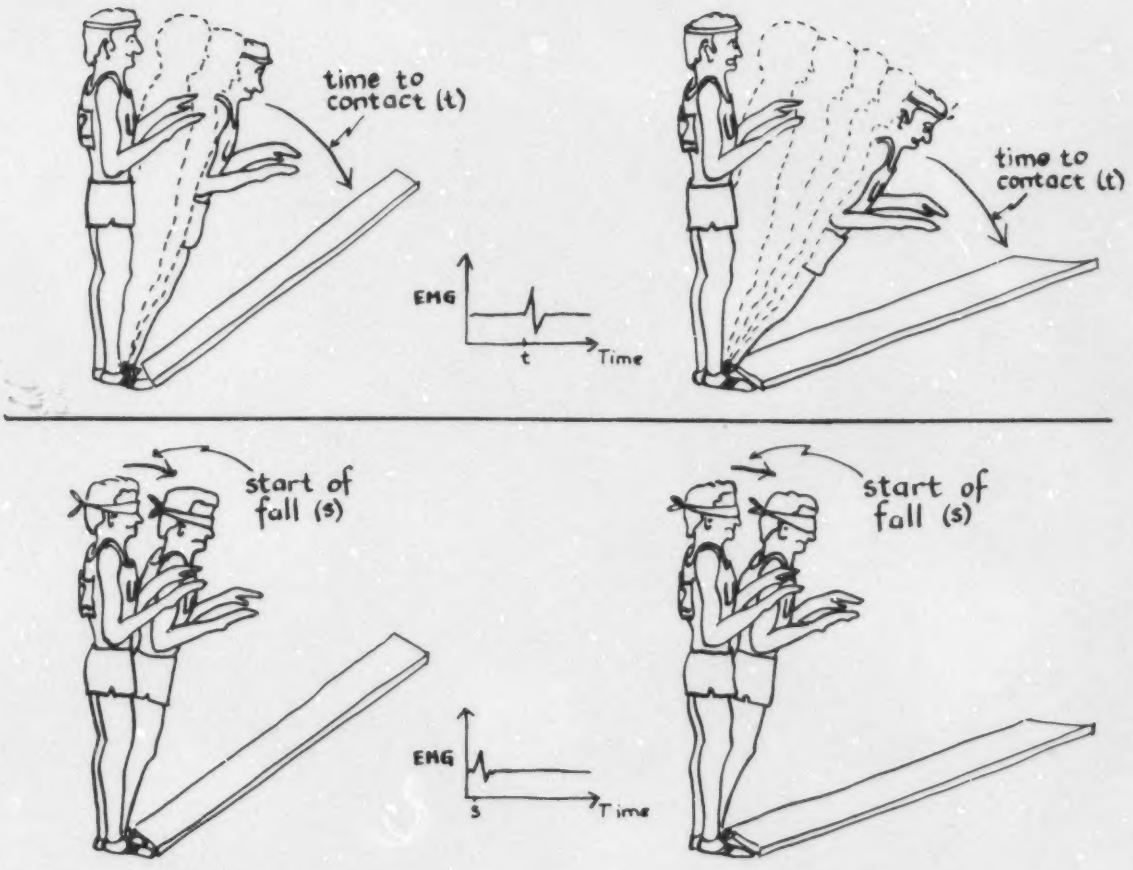


Figure 2. With the eyes open and with different falling distances the initiation of tension in the triceps brachii muscles occurs at varying times after the start of falling but at an approximately constant time prior to contact (Above). With the eyes closed initiation of muscle tension is tied to the start of falling (Below). (From Fitch, H. L., Tuller, B., & Turvey, M. T. (1982). The Bernstein perspective: III. Tuning of coordinative structures with special reference to perception. In J. A. S. Kelso (Ed.), Human motor behavior (p. 278). Hillsdale, NJ: Erlbaum.)

son, 1979; Turvey, 1977). Here we wish to comment only on the questionable strategy of analyzing higher-order activity-relevant variables in terms of the putatively more basic variables of displacement and time. Inertial guidance systems are based on Newton's laws of inertia and gravity. These systems detect accelerative forces directly. They determine velocity and distance indirectly through the single and double integration, respectively, of the accelerative forces. In like fashion, adherents to the Gibsonian program (Lee, 1980; Runeson, 1977; Turvey & Shaw, 1979) argue that the imminence of collision is not inferred from a preliminary determination of speed of approach and distance from surface; rather, the basis for an animal's knowing when a surface will be contacted is the detection of $\tau(t)$ as such. The point is that to understand how perception controls activity, we must be willing (i) to question the primary reality status of the basic variables of physics; and (ii) to look for variables (observables, quantitities) at the ecological scale that uniquely specify the relation of animal to environment; and (iii) to consider hard- or soft-molded processes that detect these ecological variables (rather than knowledge-based procedures that construct representations of them from conventional physical variables).

So, how does the animal know from where to leap? The answer, to be blunt, is that it does not need to know the proper place; rather, it needs to know the proper time. The former depends explicitly on the speed, the latter does not. Evidently, as anticipated, the successful leaping of a barrier depends on the time-to-contact variable. It also depends on body-scaled information, but we will have more to say about that below. And how does the animal know whether it is braking sufficiently? An animal's deceleration is adequate if and only if the distance it will take the animal to stop is less than or equal to its current distance from the brink (Lee, 1980). Adequacy of braking is specified by whether the rate of change of $\tau(t)$ equals or exceeds a critical value (Lee, 1976; 1980). A related observation is that flies begin to decelerate prior to contact with a surface at a critical value of $\tau(t)^{-1}$ (Wagner, 1982).

2.2.3 How does the animal know that the barrier is jumpable and that the brink is a step-down place (rather than a falling-off place)?

Knowing that something is in the class of jumpable objects and that some other thing is in the class of step-down places would be treated in the Cartesian program as the imposition of subjective, meaningful categories on an objective, meaningless surround. Conventionally, it would be said that the animal has concepts of such things and debate would focus on how such mental entities could be established. Careful analysis would reveal that, given the departure point of the Cartesian program, empirical contributions to such concepts would have to be secondary to the rational contribution (Fodor, 1975). In sharp contrast, the Gibsonian program seeks to uncover a natural, lawful basis for knowing what activity (or activities) a situation offers. Consider a brink in the surface that happens to be a step-down place for a given animal rather than a place where it would have to jump down or climb down or steer away from. To begin with, the property of the brink as a step-down place for the animal cannot be captured in the scales and standard units of physics. These scales and units are intended to be "fully objective," that is, observer- or user-independent. They are extrinsic measures, in that the standards on which they are based are divorced from and external to the situations to which they are applied. To capture a step-down place for a given animal requires intrinsic measures, those whose standards

are to be found in the situation of animal and brink. In Figure 3 the separation of surfaces (R) must somehow be expressed in units of the animal. Leg length is obviously significant but scaling surface magnitudes in terms of the unit 'eye-height' is probably a better move (cf. Lee, 1974, 1980). A lawful allometric relation (Gunther, 1975; Huxley, 1932; Rosen, 1967) is to be expected between eye height (E) and leg length (L): $L = aE^b$, where a and b are constants. (Eye height will, of course, vary with the animal's posture, but our intent here is to convey the style of the analysis rather than its full detail). If the separation of surfaces (R) at a brink is below some critical number, nE (or is less than or within a tolerance range $nE + \delta$), then the separation is a step-down place; above this critical number (or range) it is a place that requires some locomotory strategy other than stepping down. Noting that E is unity, there is a dimensionless quantity that marks the boundary between the activities--stepping down vs. jumping down, for example--that a brink offers an animal. Now the question becomes whether or not there is an optical property specific to this dimensionles quantity.

First, a point of observation moving toward a brink in a surface (where one surface partially occludes another) will lawfully generate an optical flow pattern in which there is a discontinuity, viz., a horizontal contour above which optical structure magnifies and gains and below which optical structure magnifies but does not gain. The non-gain and gain of structure are specific, respectively, to the occluding surface currently supporting the animal and the occluded surface to which it is heading. Second, from Figure 4 (after Warren, 1982) it can be seen that the separation (R) of the occluding and occluded surfaces can be expressed in units of the height of the point of observation E and in terms of the ratio of the rate of displacement of the point of observation (dx/dt) to the rate of gain of structure (ds/dt). Letting $t_1 - t_0$ be the time of a step (x), for the same stepping rate the gain of structure (s) is greater the greater the separation of surfaces (R). Although the example is crudely developed, it makes an in principle argument that there will be a dimensionless quantity of optical flow, such as (dx/dt)/(ds/dt), which is specific to the vertical separation of surfaces at a brink, scaled in units of the observer. A critical value or range of this optical flow property will specify the boundary between places the animal can step down from--those that can be accommodated by limb extension--and places requiring a different maneuver.

Dimensionless numbers play a significant role in many branches of physics. The commonly used numbers, referred to as principal numbers by Schuring (1980) (of which the Reynolds, Raleigh, Mach, Prandtl, and Froude are prime examples), are built from laws. Thus, the Reynolds number, which applies to fluids, is built from Newton's law of inertia and the law for shear stress of a Newtonian fluid. The two laws are cast as dimensionless ratios or π numbers (e.g., $\pi = F/ma$), and these two numbers in ratio give the Reynolds number. At a critical value of the Reynolds number, the inertial forces (favoring turbulence) dominate the viscous forces (favoring laminar flow) and there is a shift from the one kind of flow to the other. Generally speaking, the major dimensionless numbers in physics mark off, at critical values, a change in the relation of forces from a balance between them to a dominance by one of them and, thereby, mark off distinct physical states. In like fashion, it seems that the dimensionless numbers built from purely optical variables mark off, at critical values, distinct states. They are not physical states associated with distinct forms of energy absorption, however, but <u>specificational</u> states (see Kugler, Turvey, Carello, & Shaw, in press).

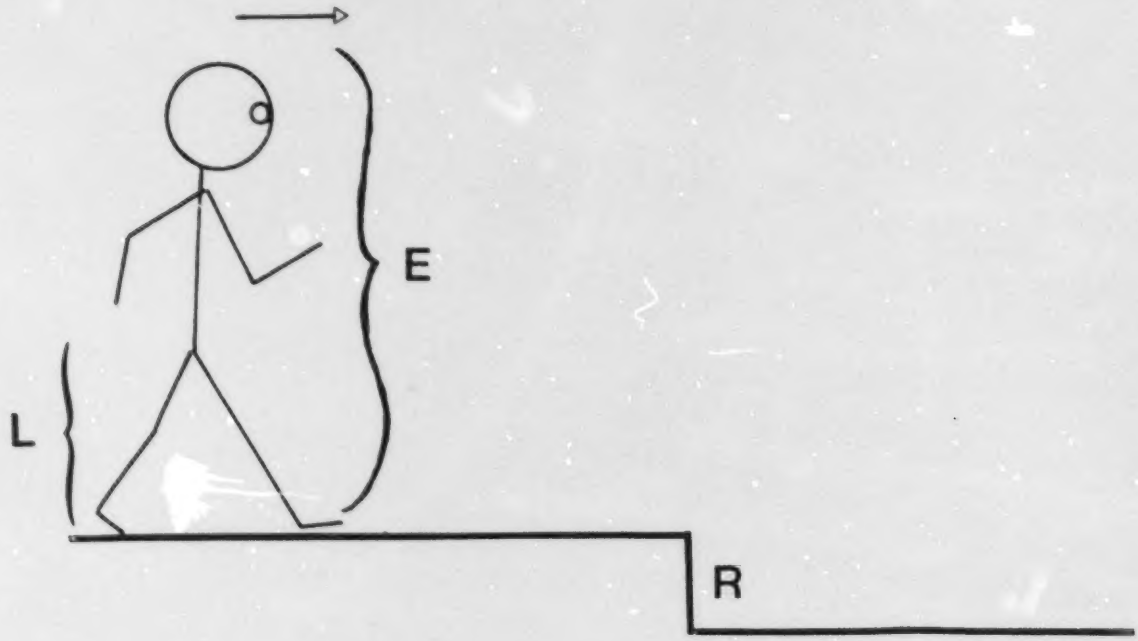


Figure 3. Approaching a brink of a surface. E is eye height, L is leg length and R is the surface separation.

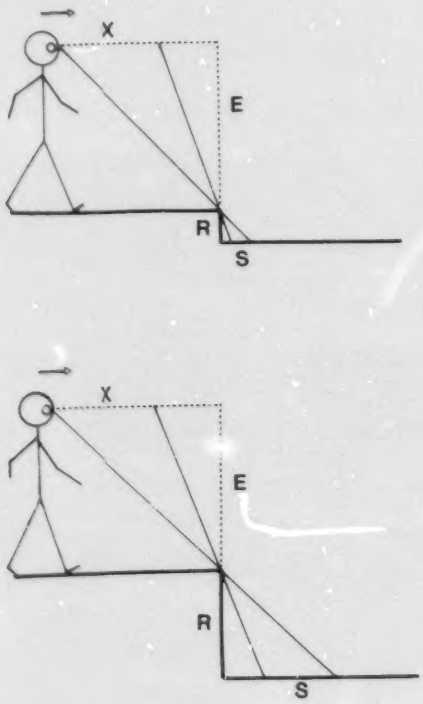


Figure 4. In approaching a brink of a surface the ratio of dX/dt/dS/dt depends on the surface separation R relative to the eye height E.

Thus, in the example just given, the dimensionless quantity (dx/dt)/(ds/dt) specifies step-downable when it is below a critical value and non-step-downable when it is above that critical value.

We wish to underscore with two well-developed examples the potential significance of dimensionless quantities to law-based explanations of the control of activity. When $d\tau(t)/dt \geq -0.5$, it specifies that the point of observation will stop prior to contact with an upcoming substantial surface if current conditions persist, whereas when $d\tau(t)/dt < -0.5$, it specifies that there will be a collision between the point of observation and the surface if the current conditions persist. This critical value of the rate of change of the time-to-contact variable is an invariant optical quantity: Whether the animal is approaching a surface or being approached by a surface, the quantity -0.5 marks off two distinct specificational states concerning the collisional consequences of the animal's current activity.

The second example returns the focus of this subsection to the perception of the kind of activity that an arrangement of surfaces affords an animal. Warren (1982) investigated the perception of stairways that varied in riser height in terms of two questions: (1) Could a person perceive whether a stairway was climbable in the normal fashion (a question of the critical riser height)? and (2) Could a person perceive how costly, in metabolic terms, a stairway would be to climb (a question of the optimal riser height)? A preliminary analysis of the biomechanics of stepping up revealed that the riser height (R) beyond which normal stair climbing would be impossible was a constant proportion of leg length (L), viz., .88L, or R/L (a dimensionless quantity) = .88. Subjects, who differed markedly in height (5'4" vs. 6'4"), saw photographs of stairways with risers that ranged between 20 in and 40 in and were asked to judge the climbableness of each stairway. Although the riser height that distinguished the stairways judged to be climbable from the stairways judged to be nonclimbable differed between the two groups of subjects when measured in inches, it did not differ when measured in leg length. For both groups of subjects R/L = .88, that is, the critical riser value that had been determined from biomechanical considerations. With respect to the optimal riser height, the metabolic cost of climbing at 50 steps/min on an adjustable, motor-driven stairmill was evaluated at riser heights varying from 5 to 10 in for short (5'4") and tall (6'4") subjects. The minimum energy expenditure per vertical meter (cal/kg-m), indicating optimal riser height, occurred at a riser height of R = .26L. In two visual tasks, a forced choice task and a rating task, the stairways were pitted against each other in pairs. The tasks revealed that the preferred riser height (the stairway that was seen to be the one that could be climbed most comfortably) differed between the two groups of subjects when measured in inches but it did not differ when measured in leg length. The preferred or optimal value for both groups was .25L in the forced choice task and .24L in the rating task, very close to the optimal value of .26L determined by metabolic measurement.

2.2.4 Affordances

In Gibson's (1979) terminology, step-down places, falling-off places, climbable-places, collide-withable surfaces, travel-throughable openings and so on (Figure 1) are affordances. That is to say, they are properties of the environment taken with reference to the animal. An affordance is an invariant arrangement of surface/substance properties that permits a given animal a particular activity. It is a real property--one might even say a physical

property--but one that is defined at the ecological scale of animals and their niches. By the laws of ecological optics, the light structured by an affordance will be specific to the affordance--as the above examples suggest. The optical property specific to an affordance is like the time-to-contact property: It is not decomposable into optical variables of a putatively more basic type. Consequently, it is claimed, the perceiving of an affordance is based on detecting the optical property that specifies the affordance. In the Gibsonian program, perceiving an affordance is not mediated by computational/representational processes. It is said to be direct, and understanding how this can be--understanding the physical processes at the ecological scale that make possible the direct perception of the reality that bears on the control of activity--is what the Gibsonian program is fundamentally about (Section 3.2).

3.0 Principles of Self-Regulation

It is fair to say that in working under the Cartesian program one is inclined to explain regularity (of activity) by reference to intelligent regulators. In the Cartesian view of things, it is an act of the intellect that interprets the outputs of sensory transducers and puts them to use with respect to externally oriented desires. Intelligence in its various manifestations (e.g., judging, comprehending, decision making, comparing, projecting and evaluating hypotheses, recognizing, reconsidering, commanding, and so on) is at the core of the Cartesian explanation of the control and coordination of movement. For Descartes himself the intellect was equated with the soul--or as Ryle (1949) liked to say, disdainfully, "the ghost in the machine."

The contemporary student of movement who chides all 'little man in the brain' explanations of control may, however, be firm in the belief that concepts borrowed from cybernetics and formal machine theory are acceptable explanatory tools. Personally, we think such convictions are suspect. Concepts such as set-points, programs, and so on are superficially attractive in that they refer to material things that perform the role historically ascribed to homunculi. Under closer scrutiny, such concepts are revealed to be the products of an intelligent act performed by a being with foreknowledge of the regularity to be achieved. The concepts of cybernetics and formal machine theory are seductive because they facilitate the simulation of 'regularities'; but they are not, we believe, in the best interests of explanatory science. First, these concepts necessarily assume intelligence and rationality--assumptions that were, after all, the reason for science's original and persistent displeasure with Descartes' homunculus. Second, their promise is limited, at best, to describing and, perhaps, to predicting regularities. But explaining, in the sense of identifying the lawful basis for behavior, is ineffably beyond their reach.

At one time, Bernstein was enthusiastic about the relevance of cybernetical and formal machine analogues to the physiology of activity. He later became much more circumspect with regard to their appropriateness. Cybernetical notions figure prominently in his discussion of "Some Emergent Problems of the Regulation of Motor Acts" (as we will underscore in the subsections that follow). But in later chapters he questions the propriety of cybernetics for biology and physiology and intimates that "the "honeymoon" between these two sciences" (p. 181) may be over (also pp. 185-186). In Section 3.1 we critically evaluate the cybernetical treatment of Bernstein's regulatory notion of circular causality and in Section 3.2 we outline the physical conditions for

that principle. Our belief, consonant with Bernstein's later impressions, is that the physiology of activity would fare better married to a physics that addresses the ecological scale and its natural regularities rather than to a formal theory of the regulation of artifacts.

3.1 The Ring Principle (Cybernetically Interpreted)

Bernstein is convinced, and properly so, that self-regulation is based on circular causality--the "ring principle" as he terms it. He embraces the familiar interpretation of this principle, the one advanced by cybernetics: A referent signal or set point mediates signals fed forward to and fed back from a device or process (generically referred to as 'the plant'). For the conduct of an activity a single referent--and, a fortiori, a single ring--will rarely be sufficient. Bernstein assumes an ordered sequence of referent signals. Insofar as a referent signal must predate the afferent and efferent flow that it mediates, so the order of the referent signals must largely be ascribed prefatory to the activity. In brief, Bernstein's proposal for self-regulation is the popular notion of a program. MacKay (1980) identifies the kinds of detail one might expect to find in a program for a step cycle of locomotion (Figure 5).

In addition to identifying (1) the general a priori prescriptive nature of the program, this example illustrates nicely that a program is (2) an orderly sequence of preferred quantities (e.g., 100 ms), (3) an orderly sequence of commands (e.g., stop extension) to the skeletomuscular machinery that realizes these quantities, and (4) an orderly sequence of symbol strings (the representational format for the quantities and the commands). It also illustrates a more profound feature of the program conception: (5) that rate-dependent processes--the irreversible thermodynamics and the mechanics of the skeletomuscular system--are coupled to and constrained by rate-independent structures--the symbol strings.

The centrality of the ring principle to self-regulation cannot be doubted. (The reciprocity of locomotion and global optical transformations described in Section 2.0 is one example of the principle's ubiquitous application.) What can be doubted is whether the properties identified in (1) through (5) above are necessarily entailed by the principle.

3.1.1 The concept of the referent signal

The sollwerts (required values, set points) that have been used frequently to 'explain' the stabilities of vegetative processes (thermoregulation, respiration, feeding, drinking, etc.) are more fictitious than real (e.g., Friedman & Stricker, 1976; Iberall, Weinberg, & Schindler, 1971; Mitchell, Snellen, & Atkins, 1970; Werner, 1977). The observed stable quantities of vegetative processes (e.g., human body temperature of 37 degrees centigrade) are not prescribed values or goals playing a causal role. They are, more accurately, resultant quantities, indexing a stable relation between independent processes (force systems) defined over the same state variables (Iberall, 1978; Kugler, Kelso, & Turvey, 1980, 1982; Yates, 1982b). As we like to put it, these so-called sollwerts are not a priori prescriptions for the system but a posteriori facts of the system's processes.

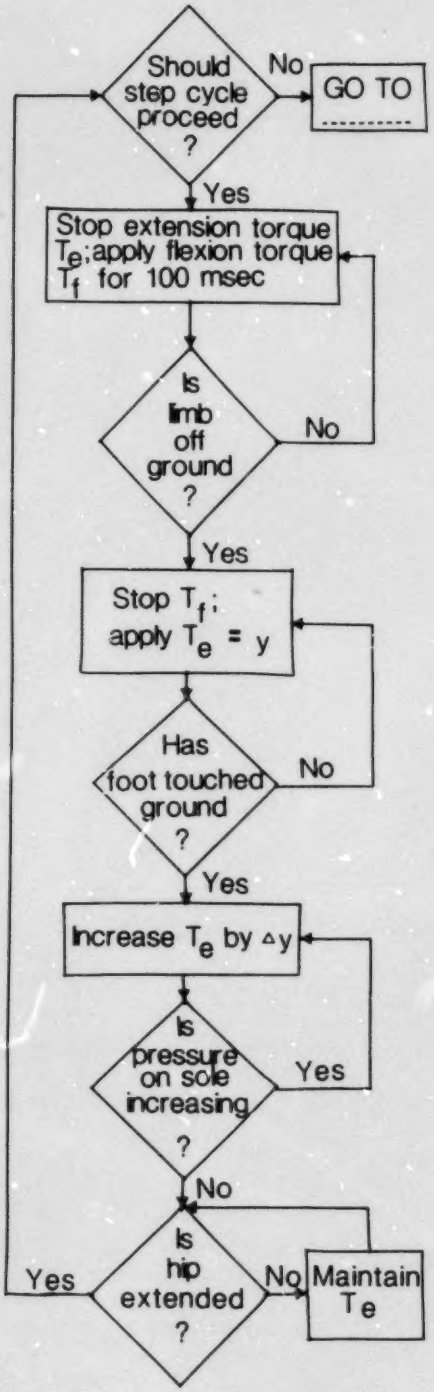


Figure 5. A program formulation of a locomotory step cycle (Adapted from MacKay, W. A. (1980). The motor program: Back to the computer. Trends in Neuroscience, 3, 97-100).

The experiments of Zavelishin and Tannenbaum (1968) are illuminating in this regard. They focussed on two respiratory variables--the resistance (r) of the air to inspiration and the duration (d) of inspiration. The function f relating d to r was identified. A function F relating r to d was imposed (Figure 6). Circular causality was thereby established. The value of d at which the respiratory system settled down was that value mutual to the two functions. If F was chosen to intersect f at more than one value of d, then the system would settle at one of the mutual points or oscillate between them depending on the actual details. Figures similar to Figure 6 are to be seen in Mitchell et al. (1970) and Werner (1977) with reference to temperature regulation, and in Guyton (1981) and Yates (1982a) with reference to the pressure-flow relationships for blood circulation.

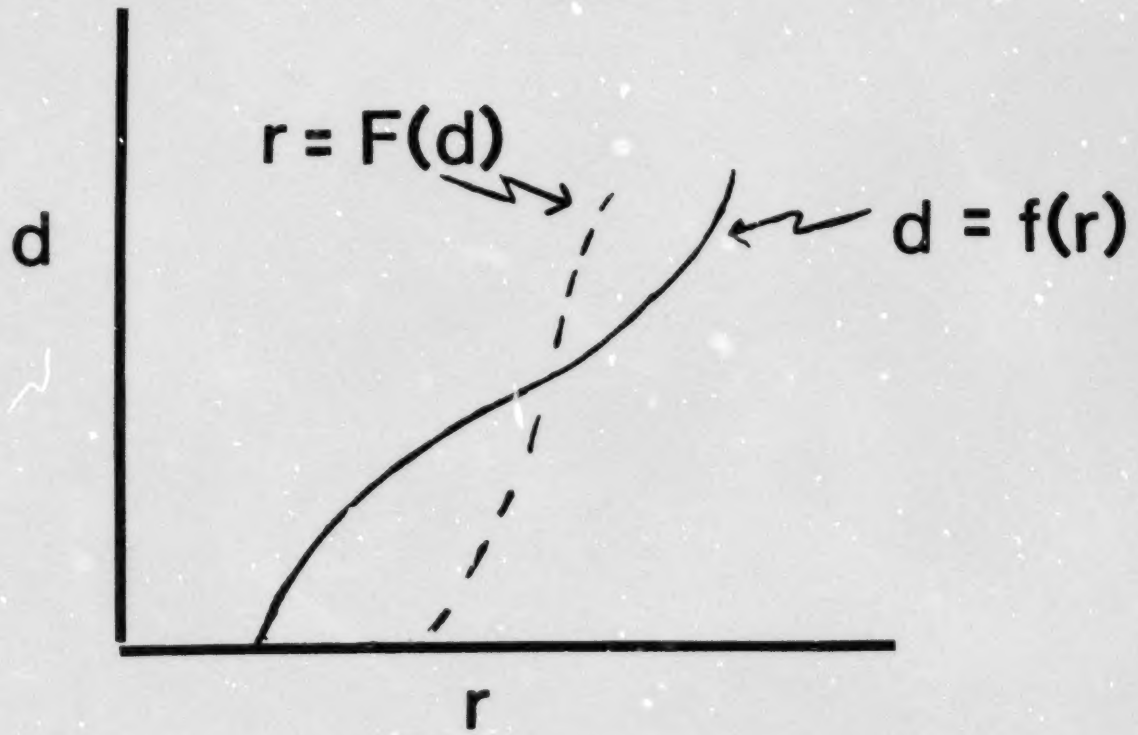


Figure 6. Circular causality defined over the respiratory variables r (resistance to inspiration) and d (duration of inspiration).

Each of the aforementioned instances of the ring principle or circular causality involves two distinct pathways of influence between two variables, x and y. The system in question must satisfy two independent causal laws, one linking x to y and one linking y to x. The real equivalent of the point of intersection of the two functions in the x by y coordinate space is the equilibrium operating point of the system--the only point that satisfies both causal laws. The equilibrium point is not frozen; it can be shifted by changing the system parameters (see Guyton, 1981; Mitchell et al., 1970; Werner, 1977; Zavelishin & Tannenbaum, 1968). In sum, each of these instances of circular causality exemplifies a stable equilibrium state that is achieved without the processes of measuring the istwert of the bounded variable, comparing it to the sollwert, and amplifying the difference to bring

about an action that reduces this difference. The processes of measurement, feedback, amplification, and comparison that Bernstein takes to be the minimal requirements of self-regulation are not to be found.

3.1.2 Intensional descriptions and teleological explanation

How general is an interpretation of self-regulation that does not implicate the conventional, intelligence-based, mediating mechanisms of cybernetics? Intuitively, a notion such as a referent signal or program, and the related processes of feedback and the like seem to be called for whenever we construct a description of a system (S) such as:

'S prefers (wants, desires, seeks, etc.) G,'

where G can be the value of a property of S, a property of a thing in S's environment, an orientation of S to the layout of the environment, and so on. A statement of the above kind is called an intensional context or description. Basically, it involves borrowing the property of one thing, G, to build a property of another thing, S, viz. 'prefers G.' What is the status of the borrowed property G?

Orthodoxy invariably interprets intensional description as license to ascribe concepts: To predicate of S the property 'prefers (wants, desires, seeks, etc.) G' is to ascribe to S the concept of G. Similarly, in matters of perception, to say that 'S can perceive step-down places' (Section 2.2.3) is to say, by the orthodox interpretation, that S has a concept of step-down places. What is it about intensional description that invites conceptual ascription? Why should the convenience of describing a property of a thing S in terms of a thing G be translated into the claim that S possesses or embodies in some form the thing G? Empirical considerations reveal that the intensional context 'biological system S prefers a body temperature of 37 degrees centigrade' does not mean that the end-state of 37 degrees is encoded in the system's central nervous system and, relatedly, does not identify a relation between S and a central nervous system representation of the quantity 37 degrees centigrade. The lesson of this example is twofold: First, intensional description does not mandate conceptual ascription; and second, intensional description may simply be a way of referring indirectly to lawful processes. It seems, therefore, that intensional description will invite conceptual ascription to the degree that a lawful basis for a given regularity is unexplored or indiscernible (Turvey et al., 1981).

Let us turn away from vegetative processes, such as temperature regulation, to the more general case. Consider the following example of a goal-directed activity, to be designated L. A swiftly flying bird suddenly changes its posture, spreads its wings, flaps them briefly, glides, flaps its wings a little more, and alights gently on a branch. A teleological description (Woodfield, 1976) of L reads:

'S did B in order to do G',

where S refers to the bird, B to the behavior and G to the goal of alighting on the branch with a minimal mutual transfer of momentum between bird and branch (Kugler, Turvey, Carello, & Shaw, in press). This teleological description of L can be expanded (after Woodfield, 1976) to make the implied internal conditions transparent, albeit in "mentalese" (Fodor, 1975):

> 'S did B because S (i) wanted to do G and (ii) believed that B would lead to G.'

Now we have a teleological explanation of L.

It is very important to distinguish a ring principle (or circular causality) explanation of L from a teleological explanation of L. The ring principle takes G for granted and explains how S gets to G. By definition, a ring principle explanation consists not so much of a single sentence of the type 'S did B (at a particular time, in a particular way, etc.) because...' but consists, rather, of a set of sentences describing cycles of acting-perceiving-change (in the animal-environment relation). The ring principle explanation of L would be in terms of the reciprocity of the bird's approach and optical flow, with particular emphasis on the decelerative forces supplied by the bird with respect to maintaining the optical flow property $d\tau(t)/dt$ within the range specifying a "soft" contact (see Section 2.2.2). The teleological explanation of L also takes G for granted, but it explains why S does B. Thus, the two explanations are complementary (Woodfield, 1976). If the processes governed by the ring principle are viewed as dynamical, then the states of S ('wanted', 'believed') are tantamount to field boundary conditions on dynamical processes. The form of these particular boundary conditions is that of non-holonomic constraints (about which more will be said below).

Clearly, the intention (i) in the teleological explanation of L above is Bernstein's 'image of achievement' (Pribram, 1971) that constrains the variations in the content of the belief (ii) until G is done. Recall that, for Bernstein, where actions are planned they are planned in terms of biological consequences (that is, in terms of how an activity will change the animal-environment relation) and not in terms of the pattern of bodily movement (Problem 3 of the introduction). But can the state picked out by the phrase 'wanted to do G' be interpreted as an 'image' in the sense of an actually existing mental or neural representation of G? Woodfield (1976) cautions thusly:

> It is tempting to think of a goal as a concrete future event, and to think of the present desire as involving a conception of that future event, with the conception of the goal being in some sense logically or ontologically derivative from the goal itself. But this is the wrong way round. A goal just is the intentional object of the relevant kind of conception. (p. 205)

Let us see what a Gibsonian analysis of G looks like. The goal G in the goal-directed activity L involves two aspects: one, a surface X that can support S and two, a soft, feet-first collision with X. The former aspect defines an affordance, and under the Gibsonian program an affordance is optically specified. That is, the light structured by a branch is specific to the support property of that surface layout vis-a-vis the bird's proportions. The latter aspect, that of the soft collision, is specified by $d\tau(t)/dt \geq -0.5$. The two aspects of the intention 'wanted to do G' in L might be interpreted in the Gibsonian program as follows: 'wanted to do G' is a matter of having detected information that continuously specified a surface of support and having detected information that continuously specified the intensity of an up-coming collision with that surface, on the occasion of a certain metabolic condition of S.

One should be circumspect about the generalizability of an analysis of the preceding type. Intentionality is a large issue and the reader's favorite example of intentional behavior is probably much more elaborate than L. However, states of affairs such as L are common; they comprise the larger part of an animal's daily directed activities. And insofar as the Gibsonian program can anchor teleological explanations of goal-directed activities such as L in natural laws at the ecological scale, it promises a natural basis for intentionality. Be that as it may, the comparison between the Cartesian and Gibsonian programs on the subject of intentional objects (the goals of goal-directed activities) is sharply drawn. Under the Cartesian program, intentional objects are represented in an internal medium; under the Gibsonian program, intentional objects are lawfully specified by structured energy distributions (Turvey et al., 1981).

3.1.3 Commands as information in the indicational sense

Any disquiet with the concepts of internally encoded required values or intentional objects as representations, extends to the concept of 'commands.' Is circular causality in general, and the perception-action ring in particular, mediated by commands? Although it has been a commonplace to say that the brain commands the body, this way of talking has been subject to little scrutiny. As Reed (1981) has observed, there is an entire theory of action wrapped up in the notion of central nervous system commands and much conceptual effort will be required to unravel it. We will give some hints of what is involved. To anticipate, issues raised in the preceding subsections will make a repeat appearance but in a subtly different form.

The control of activity is founded on information, as both Bernstein and Gibson have sought to understand. "Commands" are a kind of information that can be termed indicational because their role is to indicate an action to be performed (Reed, 1981), much as a stop sign on the highway indicates the action of arresting the forward motion of a car and a directional sign on the highway indicates which turn to take. Indicational information is incomplete. To be commanded to stop one's car is not to be told the details of how to do so. Obviously, the informational basis for controlling activity is not exhausted by information in the indicational sense. To stop the car requires information about when to begin decelerating and information about when the deceleration is sufficient and so forth. This sense of information was discussed in Section 2.2 and in the immediately preceding section. Consonant with the terminology of these earlier sections, we will refer to this sense of information as specificational. The important point to be made is that an indicated act cannot be performed without information in the specificational sense. On generalizing, this point reads: The indicational sense of information is always predicated on the specificational sense of information.

Holding this dependency in abeyance for the present, let us focus on the commonalities between commands--as sources of information in the indicational sense--and rules. Neither commands nor rules can determine an action, both commands and rules can be violated or ignored, both commands and rules can enter into conflict (creating demands for impossible outcomes), and both commands and rules require an explicit act of comprehension for their functioning (Reed, 1981). For these reasons, a lawful determinate account of the control and coordination of activity cannot be founded on the notion of commands or information in the indicational sense. A further undesirable feature is that the criticisms that apply to a body-states or sensation-based

theory of perception (see Section 2.0) apply to a command-based theory of action: There is no rational explanation of the genesis of the knowledge that forms and interprets commands. A command-based theory of action looks like another unrepayable loan of intelligence.

The lawful basis of optical structure relevant to activity's control was labored in Section 2.2 in order to make the notion of specification transparent. Where information in the indicational sense is close to the concept of rule, information in the specificational sense is cognate with law. Laws are determinate, non-negotiable (they can never be violated or ignored), harmonious (they can never give rise to impossibilities), and they do not depend on explicit knowledge for their functioning. In the cybernetical interpretation, the ring-principle is mediated by indicators (commands). But it is apparent that this need not be so, for the same reason that mediation by reverent signals need not be so. It is an unmediated, law-based interpretation of the ring principle (rather than a mediated, rule-based interpretation) that is the focus of the Gibsonian program (see Section 3.2). A lawful account of the control and coordination of activity cannot be founded on information in the indicational sense but it could be founded on information in the specificational sense.

3.1.4 Symbolic and dynamical modes

The contrast of indicational information and specificational information parallels that of discrete symbol strings and continuous dynamical processes or, equivalently, rate-independent structures and rate-dependent processes. These contrasts are said by Pattee (1973, 1977, 1979) to identify a Complementarity Principle that is the hallmark of living systems. Living systems are seen to execute in two modes, the symbolic and the dynamic, which are incompatible and irreducible. Consequently, understanding biological, physiological and psychological phenomena is said by Pattee to rest with the elaboration of this complementarity. The computational/representational approach to these phenomena that is championed by the Cartesian program is flawed--in Pattee's view--because it attempts to explain only through the discrete symbolic mode. Similarly, in his view, an approach that seeks to explain such phenomena using only (sic) the laws of dynamics will also prove inadequate. By Pattee's reasoning, both modes must be given full recognition; the phenomena in question are the result of the coordination between the two modes. Stated more sharply, complementarity is advanced as a principle that calls for simultaneous use of formally incompatible descriptive modes in the explanation of the characteristic phenomena of living systems (Pattee, 1982).

There is, however, an asymmetry between the two modes that has to be appreciated. Nature uses the symbolic mode--nonholonomic (nonintegrable) constraints--sparingly. Dynamics are used to the fullest, wherever and whenever, to achieve characteristic biological effects. Symbol strings are used, now and then, to direct dynamical processes and to keep down their complexity--in other words, to trim the dynamical degrees of freedom. In Figure 5, which depicts a prototypical program formulation of activity, the opposite strategy is at work. Very many nonholonomic constraints are exploited to achieve ('to explain') the kinetic and kinematic regularities of a locomotory step cycle. The question of how the dynamics--properly construed for the biological scale in terms of the conjunction of statistical mechanics and irreversible thermodynamics (Iberall, 1977; Prigogine, 1980; Soodak & Iberall, 1978)--might fashion the phenomenon is not addressed, nor is the question of

how the symbol strings interface with the dynamics. Pattee's analysis is an important one for those students of movement who would pursue the Cartesian program with its emphasis on the symbolic mode: Only in the working out of the physics of a regularity can one identify the nature and type of symbol strings (nonholonomic constraints) needed to complete the explanation. To begin with the symbolic mode, and to adhere strictly to it, invites an account that will be plagued by arbitrariness (as, surely, is the account of a step-cycle represented by Figure 5). To begin with the dynamical mode, and to pursue it earnestly, promises an account that will be principled.

There is, however, a deep problem with Pattee's Complementarity Principle. For Pattee, the discrete symbol strings function as information in the indicational sense. The proposed complementarity, therefore, is one of indicational information and dynamics. The problem with endorsing a view of indicational information and dynamics as formally incompatible is that it rules out any explanation of the origin of indicational information. We and others have recorded our disquiet with the Complementarity Principle for just this reason (Carello, Turvey, Kugler, & Shaw, 1982; Kugler et al., 1982). One suspects that for the consistency of physical theory, information in the indicational sense should be lawfully derivable from dynamics (or information in the specificational sense).

3.2 The Ring Principle (Physically Interpreted)

In this final section we provide an overview of the physical foundations of Bernstein's ring principle. Whereas the cybernetical interpretation of the ring principle is consistent with the Cartesian program, the interpretation evolving in physical biology is consistent with the Gibsonian program.

3.2.1 Open systems and the role of causal dynamics

According to classical physics, living systems are continuously struggling against the laws of physics. Within the last few decades, however, it has become increasingly apparent that those physical systems that are open to the flow of energy and matter into and out of their operational components behave in a manner that suggests the behavior of living things and suggests a dramatically different view of causal dynamics (see Yates, 1982a, 1982b, for a review). Whereas the behavior of an isolated physical system is strictly determined by the system's initial and boundary conditions, systems open to the flow of energy and matter can evolve internal constraints that 'free' the system's dynamics from its initial conditions. The arising of the new internal constraints serves to limit the trajectories of the internal components, thereby reducing the system's internal degrees of freedom. As these constraints arise, new spatio/temporal orderings are created and the system derives new ways of doing business with its surroundings (that is, new ways of transacting energy).

While living systems can be viewed as following from the laws of physics, one distinguishing characteristic that emerges in systems of this order of complexity is the ability to time-delay energy flows internally. This is accomplished through the maintenance of internal potentials from which the system can periodically draw energy so as to produce a generalized external work cycle. This self-contained source of potential energy (usually in chemical form) allows the system to be characterized as self-sustaining. The ability to be self-sustaining means that the system's behavior is no longer governed

strictly by minimum energy trajectories or external work cycles defined on surrounding gradient fields. The possibility now arises that a self-sustained system can temporarily depart from the constraints defined by the surrounding potential minimums. Departures from and returns to minimum regions defined in the surrounding potential field require some form of sensitivity to the gradients; and this, in turn, requires some form of self-sustaining system. The ability to discriminate low order potential gradients selectively (Frohlich, 1974; Volkenstein & Chernavskii, 1978) and the ability to form an autonomous, persistent self-sustaining system (Iberall, 1972, 1977) are fundamental characteristics of living systems.

3.2.2 Determinate and nondeterminate trajectories in particle/field systems

Particle physics (classical, quantum and relativistic) studies the trajectories of particles to infer the dispositions of potential fields. The assumption underlying the above strategy is that variations in the observed force field are strictly a function of the particle's position in the field. The above assumption rests on two requirements: (i) that surrounding potentials remain constant (in both space and time), and (ii) that the particle has no internal means for introducing or absorbing forces (which could contribute to a trajectory's departing from the minimum regions defined by the surrounding potential field). Particles satisfying these two requirements have their trajectories completely determined by the form of the surrounding potential field: The minimum regions identify geometrical singularities in a topological field. The particle system is completely determined by and causally dependent on the topological form of the surrounding potential field.

3.2.3 Self-sustaining systems and circular causality

If, however, the particle system of interest has an internal means for generating and dissipating forces of a magnitude comparable to the forces of the surround--that is, the system is self-sustaining--then the behavior of the particle need not be completely determined by the topological form of the surrounding potential field. The particle has available internal sources and sinks that can generate and absorb forces that, when combined with the forces generated by the surrounding potential field, can yield equilibrium states that are not strictly defined by the singularities of the surrounding potential field. The behavior of this class of particle system can be said to be nondeterminate with respect to its relationship with the surrounding potential field (cf. Kugler, Kelso, & Turvey, 1982; Kugler, Turvey, & Shaw, 1982). While the particle's equilibrium states are no longer determinately specified by the state of the surrounding potential field, the particle is still, nonetheless, causally coupled to the forces generated by the surrounding potential. That is to say, changes in the forces generated by the surrounding potential field will require compensatory changes in the particle's internally generated forces if an equilibrium state is to be maintained invariant: The forces generated by the surround and the forces generated internally are causally linked in a circular causality with respect to invariant equilibrium states.

The physical concept of circular causality (cf. Iberall, 1977) is meant to identify the lawful nature of the coupling that links the surrounding potential field (and its associated force field) with the interior potential field (and its associated force field). Self-sustaining systems and their associated equilibrium states are lawfully coupled to the surrounding poten-

tial field through circular causality; they are systems whose interior potential fields play an active role in fashioning final equilibrium states.

Self-sustaining particle systems are characterized by low energy couplings that relate the particle's position to its surrounding field. The low energy coupling is defined relative to the external work cycle generated by the particle. The coupling defines a ratio of the forces generated by the particle's external work cycle in proportion to the forces generated by the surrounding potential field. A dimensionless number can be used to distinguish the nature of the coupling qualitatively:

$$Pi = \frac{\text{(forces generated by the particle's external work cycle)}}{\text{(forces generated by the surrounding potential field)}}$$

$Pi \leq 1$ = high energy coupling

$Pi > 1$ = low energy coupling

A high energy coupling ($Pi \leq 1$) defines a coupling in which the particle's external work cycle is such that it is insufficient to resist the surrounding field's potential gradients. If, however, an external work cycle is generated by the particle that resists the surrounding field's potential gradients ($Pi > 1$), and contributes actively in the organization of equilibrium states, then the coupling can be considered to be of a low energy nature. The low energy coupling realized by a self-sustaining system forms a lawful basis from which a generalized theory of information can be derived.

3.2.4 Information and the ecological approach to perception and action

Central to the Gibsonian program is the claim that information must refer to physical states of affairs that are specific and meaningful to the control and coordination requirements of activity (Turvey & Carello, 1981).

Following Gibson (1950, 1966, 1979) the above requirements for information are to be found in the qualitative properties captured in the structured patterns of energy distributions coupling an animal to its environment (see Section 2.2). These patterns (1) carry, in their topological form, properties that are specific to components of change and components of persistence in the animal-environment relation; (2) are meaningful (i.e., they define gradient values) with respect to the animal's internal potentials; and (3) are lawfully determined by the environment and by the animal's movements relative to the environment. According to the Gibsonian program, information is a physical variable that defines a coupling that is specific and meaningful with respect to the changing geometry of the econiche (defined by the animal/environment qua particle/field system totality). The energy patterns coupling the animal (internal potential field) and environment (surrounding potential field) are continuously scaled to the changing parameters and dimensionality of the system (cf. Kugler et al., 1980). The information carried in the evolving geometry of structured energy distributions is information about animal dynamics (internal potential layout) relative to environmental dynamics (surrounding potential layout). This concept of information is consistent with Thom's view of information as geometric form (cf. Kugler et al., 1982):

> any geometric form whatsoever can be the carrier of information, and in the set of geometric forms carrying information of the same type

> the topological complexity of the form is the qualitative scalar of the information. (Thom, 1975, p. 145)

Information as a geometry of form (defined over potential fields) arises as an *a posteriori* fact of the system. The information can be carried in the form of geometric manifolds that are created, sustained, and dissolved within a large variety of physical *flow fields*. The flow fields can be assembled out of mechanical, chemical, or electro-magnetic constraints.

3.2.5 On the determinate nature of information and the non-determinate nature of behavior

The goal of physics for the twentieth century has been to understand the nature of the energy states exhibited by particles at all scales of magnitude. The foundation of physics rests on the commitment (explicit or implicit) to natural laws, that is, the commitment to a natural continuity in energy states reducible to symmetry statements (equations) defined on conservations. The strategy for defining natural laws rests on the identification of trajectories assumed by particles. While this strategy has valid application for simple particle systems (non-self-sustaining systems), its application toward explicating the natural laws governing the energy states of self-sustaining systems must be questioned seriously. The behavior of a self-sustaining system is *not* strictly determined by the energy states of the surrounding potential fields. As noted, the energy states of the internal potential field play an active role in the determination of the observed trajectory. While the behavior (observed trajectory) of a self-sustaining system has a nondeterminate status, the informational states defining the low energy coupling that relates the surrounding potential field to the internal potential field has a determinate status. The information states are invariant (i.e., stable and reproducible) in the strictest sense of lawful determinism. A physical analysis of the behavior (observed trajectories) of self-sustaining systems must entail an inquiry into the low energy informational states that lawfully couple an animal (complex self-sustaining particle) to its environment (surrounding potential field). (For an example of a physical analysis of the role of low energy couplings, see Kugler, Turvey, Carello, & Shaw, in press). *It can be argued that the goal of a physics befitting Bernstein's physiology of activity is that of identifying the laws that create, sustain and dissolve low energy informational states.*

REFERENCES

Bernstein, N. A. (1967). *The coordination and regulation of movements*. London: Pergamon Press.

Boynton, R. (1975). The visual system: Environmental information. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* Vol. III. New York: Academic Press.

Carello, C., Turvey, M. T., Kugler, P. N., & Shaw, R. E. (1982). Inadequacies of the computer metaphor. *Haskins Laboratories Status Report on Speech Research*, SR-71/72, 229-249.

Chomsky, N. (1980). Rules and representations. *The Behavioral and Brain Sciences*, 3, 1-62.

Cummins, R. (1977). Programs in the explanation of behavior. *Philosophy of Science*, 44, 269-287.

Dietz, V., & Noth, J. (1978). Pre-innervation and stretch responses of triceps bracchi in man falling with and without visual control. *Brain Research*, 142, 576-579.

Fitch, H., & Turvey, M. T. (1978). On the control of activity: Some remarks from an ecological point of view. In D. Landers & R. Christina (Eds.), Psychology of motor behavior and sports. Urbana, IL: Human Kinetics.

Fitch, H. L., Tuller, B., & Turvey, M. T. (1982). The Bernstein perspective: III. Tuning of coordinative structures with special reference to perception. In J. A. S. Kelso (Ed.), Human motor behavior. Hillsdale, NJ: Erlbaum.

Fodor, J. A. (1975). The language of thought. New York: Thomas Y. Crowell.

Fodor, J. A. (1980). Methodological solipsim considered as a research strategy in cognitive psychology. Behavioral and Brain Science, 3, 63-109.

Fodor, J. A., & Pylyshyn, Z. (1981). How direct is visual perception? Some reflections on Gibson's 'Ecological Approach.' Cognition, 9, 139-196.

Freedman, W., Wannstedt, G., & Herman, R. (1976). EMG patterns and forces developed during step-down. American Journal of Physical Medicine, 55, 275-290.

Friedman, M. I., & Stricker, E. M. (1976). The physiological psychology of hunger: A physiological perspective. Psychological Review, 86, 409-431.

Frohlich, H. (1974). Collective phenomena of biological systems. In H. Haken (Ed.), Cooperative effects: Progress in synergetics. Amsterdam: North Holland.

Gelfand, I. M., Gurfinkel, V. S., Fomin, S. V., & Tsetlin, M. L. (Eds.) (1971). Models of the structural-functional organization of certain biological systems. Cambridge, MA: MIT Press.

Gershun, A. (1939). The light field. Journal of Mathematics and Physics, 18, 51-151.

Gibson, J. J. (1950). The perception of the visual world. Boston: Houghton-Mifflin.

Gibson, J. J. (1961). Ecological optics. Vision Research, 1, 253-262.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). The ecological approach to visual perceptions. Boston: Houghton-Mifflin.

Gibson, J. J. (1982). Notes on action. In E. Reed & R. Jones (Eds.), Reasons for realism: Selected essays of James J. Gibson. Hillsdale, NJ: Erlbaum.

Gunther, B. (1975). Dimensional analysis and theory of biological similarity. Physiological Review, 55, 659-699.

Guyton, A. C. (1981). Textbook of medical physiology. Philadelphia: W. B. Saunders.

Hanson, N. R. (1958). Patterns of discovery. Cambridge: Cambridge University Press.

Hofsten, C. von (1983). Catching skills in infancy. Journal of Experimental Psychology: Human Perception and Performance, 9, 75-85.

Hofsten, C. von, & Lindhagen, K. (1979). Observations on the development of reaching for moving objects. Journal of Experimental Child Psychology, 28, 158-173.

Holst, E. von, & Mittelstädt, H. (1950). Das Reafferenzprinzip. Naturwiss, 37, 464-476.

Hubbard, A. W., & Seng, C. N. (1954). Visual movements of batters. Research Quarterly, 25, 42-57.

Huxley, J. (1932). Problems of relative growth. London: Metheun.

Iberall, A. S. (1972). Toward a general science of viable systems. New York: McGraw-Hill.

Iberall, A. S. (1977). A field and circuit thermodynamics for integrative physiology: I. Introduction to general notions. American Journal of

Physiology/Regulatory, Integrative, & Comparative Physiology, 2, R171-R180.

Iberall, A. S. (1978). A field and circuit thermodynamics for integrative physiology: III. Keeping the books - a general experimental method. American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology, 3, R85-R97.

Iberall, A. S., Weinberg, M., & Schindler, A. (1971, June). General dynamics of the physical-chemical systems in mammals. NASA Contractor Report 1806.

Johansson, G. (1970). A letter to Gibson. Scandinavian Journal of Psychology, 11, 67-74.

Koenderink, J. J., & van Doorn, A. J. (1981). Exterospecific component of the motion parallax field. Journal of the Optical Society of America, 71, 953-957.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 1-47). New York: North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination. Chichester: John Wiley.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. (in press). The physics of controlled collisions: A reverie about locomotion. In W. H. Warren, Jr. & R. Shaw (Eds.), Persistence and change: Proceedings of the first international conference on event perception. Hillsdale, NJ: Erlbaum.

Kugler, P. N., Turvey, M. T., & Shaw, R. E. (1982). Is the "cognitive impenetrability condition" invalidated by contemporary physics. The Behavioral and Brain Sciences, 2, 303-306.

Lee, D. N. (1974). Visual information during locomotion. In R. McLeod & H. Pick (Eds.), Perception: Essays in honor of J. J. Gibson. Ithaca, NY: Cornell University Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 281-295). Amsterdam: North-Holland.

Lee, D. N., & Reddish, P. E. (1981). Plummeting gannets: A paradigm of ecological optics. Nature, 293, 293-294.

Lishman, R., & Lee, D. N. (1973). The autonomy of visual kinaesthetisis. Perception, 2, 287-294.

MacKay, W. A. (1980). The motor program: Back to the computer. Trends in Neuroscience, 3, 97-100.

Michaels, C. F., & Carello, C. (1981). Direct perception. New York: Prentice-Hall.

Mitchell, D., Snellen, J. W., & Atkins, A. R. (1970). Thermoregulation during fever: Change of set-point or change of gain. Pflügers Arch, 321, 293-302.

Moon, P. (1961). Scientific basis of illuminating engineering. New York: Dover.

Moon, P., & Spencer, D. E. (1981). The photic field. Cambridge, MA: MIT Press.

Pattee, H. H. (1973). Pnysical problems of the origin of natural controls. In A. Locker (Ed.), Biogenesis, evolution, homeostasis (pp. 41-49). Heidelberg: Springer-Verlag.

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. International Journal of General Systems, 3, 259-266.

Pattee, H. H. (1979). The complementarity principle and the origin of macromolecular information. Biosystems, 11, 217-226.

Pattee, H. H. (1982). The need for complementarity in models of cognitive behavior--A response to Carol Fowler and Michael Turvey. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes (Vol. 2). Hillsdale, NJ: Erlbaum.

Pribam, K. H. (1971). Languages of the brain. Englewood Cliffs, NJ: Prentice-Hall.

Prigogine, I. (1980). From being to becoming: Time and complexity in the physical sciences. San Francisco: W. H. Freeman & Co.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. The Behavioral and Brain Sciences, 3, 111-169.

Reed, E. S. (1981, November). Indirect action. Unpublished manuscript, Center for Research in Human Learning, University of Minnesota.

Reed, E. S. (1982a). An outline of a theory of action systems. Journal of Motor Behavior, 14, 98-134.

Reed, E. S. (1982b). The corporal idea hypothesis and the origin of experimental psychology. Review of Metaphysics, 35, 731-752.

Reed, E., & Jones, R. (1982). Reasons for realism: Selected essays of James J. Gibson. Hillsdale, NJ: Erlbaum.

Rosen, R. (1967). Optimality principles in biology. New York: Plenum Press.

Runeson, S. (1977). On the possibility of "smart" perceptual mechanisms. Scandinavian Journal of Psychology, 18, 172-179.

Ryle, G. (1949). The concept of mind. London: Hutchinson.

Schuring, D. J. (1980). Scale models in engineering. New York: Pergamon Press.

Sharp, R. H., & Whiting, H. T. A. (1974). Exposure and occluded duration effects in ball-catching skill. Journal of Motor Behavior, 6, 139-147.

Sharp, R. H., & Whiting, H. T. A. (1975). Information-processing and eye movement in a ball-catching skill. Journal of Movement Studies, 1, 124-131.

Shaw, R. E., Turvey, M. T., & Mace, W. (1983). Ecological psychology: The consequences of a commitment to realism. In W. Weimer & D. Palermo (Eds.), Cognition and the symbolic processes (II). Hillsdale, NJ: Erlbaum.

Smokler, H. E. (1968). Conflicting conceptions of confirmation. Journal of Philosophy, 115, 300-312.

Soodak, H., & Iberall, A. S. (1978). Homeokinetics: A physical science for complex systems. Science, 201, 579-582.

Srinivasan, M. V. (1977). A visually-evoked roll response in the housefly: Open-loop and closed-loop studies. Journal of Comparative Physiology, 119, 1-14.

Thom, R. (1975). In D. H. Fowler (Trans.), Structural stability and morphogenesis. Reading, MA: Benjamin, Inc.

Turvey, M. T. (1977). Contrasting orientations to the theory of visual information processing. Psychological Review, 84, 67-88.

Turvey, M. T. (1979). The thesis of efference-mediation of vision cannot be rationalized. The Behavioral and Brain Sciences, 2, 81-83.

Turvey, M. T., & Carello, C. (1981). Cognition: The view from ecological realism. Cognition, 10, 313-321.

Turvey, M. T., & Remez, R. (1979). Visual control of locomotion in animals: An overview. In Proceedings of Conference on Interrelations among the communicative senses. NSF publication.

Turvey, M. T., & Shaw, R. E. (1979). The primacy of perceiving: An ecological reformulation of perception for understanding memory. In L-G. Nilsson (Ed.), Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary. Hillsdale, NJ: Erlbaum.

Turvey, M. T., Shaw, R. E., & Mace, W. (1978). Issues in the theory of action: Degrees of freedom, coordinative structures and coalitions. In J. Requin (Ed.), Attention and performance VII. Hillsdale, NJ: Erlbaum.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). Cognition, 9, 237-304.

Volkenstein, M. V., & Chernavskii, D. S. (1978). Information and biology. Journal of Social and Biological Structures, 1, 95-108.

Wagner, H. (1982). Flow-field variables trigger landing in flies. Nature, 297, 147-148.

Warren, W. (1982). A biodynamic basis for perception and action in bipedal climbing. Unpublished doctoral dissertation, University of Connecticut.

Werner, J. (1977). Mathematical treatment of structure and function of the human thermoregulatory system. Biological Cybernetics, 25, 93-101.

Woodfield, A. (1976). Teleology. Cambridge: Cambridge University Press.

Yates, F. E. (1982a). Outline of a physical theory of physiological systems. Canadian Journal of Physiology and Pharmacology, 60, 217-248.

Yates, F. E. (1982b). Systems analysis of hormone action: Principles and strategies. In R. F. Goldberger (Ed.), Biological regulation and development Vol. III - Hormone action. New York: Plenum.

Zavelishin, N. V., & Tenenbaum, L. A. (1968). Control processes in the respiratory system. Automation and Remote Control, 9, 1456-1470.

239

MAPPING SPEECH: MORE ANALYSIS, LESS SYNTHESIS, PLEASE*

Michael Studdert-Kennedy+

Stimulation mapping would be of little interest, if its achievements were merely to assign brain loci to categories of linguistic or psychological description. Our understanding of complex, intermodal functions, such as naming or reading, would be blocked rather than advanced, if we were to conclude, as Ojemann speculates, that each is a macrocolumn or module, an impenetrable, vitreous chip in the great mosaic of language. The promise of stimulation mapping is rather that it may reveal, by some pattern of association and dissociation, the simpler mechanisms from which a function emerges and, ultimately, its underlying neural circuitry. To fulfill this promise, stimulation studies should not adopt uncritically the familiar, nonanalytic, modality-based tests of aphasia assessment. Rather, these tests or others must be given a functional analysis in terms of clear-cut psycholinguistic hypotheses. Naming, for example, is not a unitary function: naming errors may reflect perceptual, semantic, or phonological deficits (Goodglass, 1980), and the source of naming errors may often be inferred from their form (e.g., Katz, 1982). Similarly, deficits in oral reading are open to increasingly sophisticated analysis in terms of phonological segmentation, lexical access, and phonetic execution (e.g., Liberman, Liberman, Mattingly, & Shankweiler, 1980). Unfortunately, little in the target paper suggests that systematic analysis of this kind was attempted.

Ojemann did, however, test one "lower" function that might plausibly be expected to enter into a pattern of relations with several others, namely, orofacial mimicry. If the ability to produce simple movements of the mouth is impaired, one would not be surprised if the ability to speak, in naming or reading, or even to recall a word (if short-term storage engages a motor representation), were also impaired. In fact, very much these relations were observed (though not with perfect consistency). Yet interpretation of even this modest pattern of associations is hazardous.

Before turning to this, consider how we might interpret the most controversial link with impaired mimicry: impaired phoneme identification. The key, at least to the frontal lobe sites, is hinted at in the sparsely reported findings of Darwin, Taylor, and Milner (Ettlinger, Teuber, & Milner, 1975,

---

*Commentary on Ojemann, G. A. Brain organization for language from the perspective of electrical stimulation mapping. The Behavioral and Brain Sciences, 1983, 6, 218-219.

+Also Queens College and Graduate Center, City University of New York.

Acknowledgment. My thanks to Virginia Mann for advice. Preparation of this comment was supported in part by NICHD Grant No. HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

p. 132), cited by Ojemann. These authors discovered that patients in whom the left facial regions had been excised, for relief of epilepsy, were impaired both in spelling to dictation and in the same phoneme identification task as Ojemann used. Since these patients could understand and talk normally, Darwin and his colleagues concluded that their difficulty was confined to tasks stressing the phonetic structure of speech sounds. This conclusion implies that the identification of phonemes in nonsense words may be essentially a nonlinguistic task (inasmuch as it bypasses the lexical and syntactic processes of normal speech perception), a task very close, in fact, to mimicry. Ojemann stresses that stimulation during the phoneme identification test occurred only during presentation of the stimulus. However, since the patients had simply to reproduce what they heard, their task reduced to finding, during presentation, the motor pattern specified by the stimulus. Moreover, if, as Ojemann suggests, motor sequence programs were stored in the temporo-parietal region, this account would also handle the temporo-parietal links between phoneme identification (accomplished by execution of a two-syllable nonsense word) and three-gesture mimicry.

We may note, incidentally, that all phoneme identification tasks, calling for metalinguistic judgments, may reveal more about structural correspondences between audition and articulation than about normal processes of speech perception. Accordingly, even if a motor-reference theory of speech perception were still viable, Ojemann's findings would have little bearing on it. In fact, the link between perception and production is probably deeper and less tortuous than the old motor theory proposed. As Ojemann himself hints, the link is clearest in language acquisition, when the child learns to speak by discovering the articulatory dynamics specified by the speech it hears.

I come next to the pattern of associations and dissociations between functions, rather confusingly charted for a seven-subject series in Ojemann's Figures 3 and 4.[1] Apparently, the circles of Figure 3 correspond to the large circles of Figure 4. There are 52 large circles (25 frontal, 9 parietal, 18 temporal), representing sites where all five language functions were tested. In addition, Figure 4 displays (by my count) 27 small circles (4 frontal, 10 parietal, 13 temporal), representing sites where all functions, except orofacial mimicry, were tested. This brings us to a total of 79 sites. Of these, 15 seem to have yielded no result, leaving us with 64 effective sites (24 frontal, 18 parietal, 22 temporal), an average of about 9 per patient. By arduous tabulation, we can discover the links between functions impaired at each site—though not, unfortunately, how these links were distributed across subjects.

Among my findings was the fact, touched on by Ojemann in his discussion of discrete localization, that every function (except, interestingly, mimicry) is disturbed alone on at least one site, a dissociation that effectively demonstrates the absence of causal relations among the functions in at least those patients who display it. But what is most remarkable is that every function (except single-gesture mimicry, confined to the frontal lobe) is impaired in every area. Thus, short-term memory (STM) is impaired at 16/24 frontal, 11/18 parietal, 7/22 temporal sites; reading and/or naming are impaired at 13/24 frontal, 9/18 parietal, 11/22 temporal sites; three-gesture mimicry is impaired at 4/23 frontal, 5/8 parietal, 5/13 temporal sites; phoneme identification is impaired at 9/24 frontal, 5/18 parietal, and 6/22 temporal sites. How are we to square this distribution with Ojemann's model,

assigning retrieval to the frontal, storage to the temporo-parietal lobes? Moreover, Ojemann reports that STM was tested with stimulation at the time of input, storage, or retrieval, but these distinctions are not preserved in the report of the data. Were all frontal STM deficits confined to retrieval, all temporo-parietal deficits to storage?

The problem worsens as soon as we leave these statistical patterns and consider individual subjects--a reasonable move if we are interested in mechanism. Ojemann acknowledges a high degree of individual variability, but assures us that "the interrelationships described in the model can be readily identified in individual patients, such as the one illustrated in Figure 2." Why, then, were we not given a comparable figure (or at least a table) laying out the pattern of relations for each subject? Yet, even if we had the individual data, we would have to be cautious in interpretation. If two functions are dissociated, we can be confident that there is no necessary connection between them. However, even if they are regularly associated, we cannot infer a necessary connection. This is so because electrodes are large relative to nervous tissue, so that we cannot be sure that the association is not due to blocking of distinct, though closely neighboring, functions. These limitations are inherent in the stimulation-mapping technique in its present state of refinement.

On the other hand, as Ojemann argues, the fact that certain associations do recur over wide areas of peri-Sylvian cortex is encouragement enough for continued research. Ojemann is to be honored for rediscovering a valuable technique, the most precise that we have to analyze the neural circuitry of functioning human cortex, and for leading the way toward important new discoveries.

References

Ettlinger, G., Teuber, H-L., & Milner, B. (1975). The Seventeenth International Symposium of Neuropsychology. Neuropsychologia, 13, 125-134.

Goodglass, H. (1980). Disorders of naming following brain injury. American Scientist, 68, 647-655.

Katz, R. B. (1982). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. Unpublished doctoral dissertation, University of Connecticut.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia. Baltimore, MD: University Park Press.

Footnote

[1]Figures from the original paper by Ojemann not reproduced.

BOOK REVIEW*

R. E. Asher and Eugenie J. A. Henderson (Eds.). (1981). Towards a history of phonetics (pp. i-317). Scotland: Edinburgh University Press.

Leigh Lisker+

The editors of this volume, a collection of eighteen papers on various topics in the history of phonetic ideas and of writing systems, had two purposes in undertaking their enterprise: to do honor to David Abercrombie on the related occasions of his seventieth birthday and retirement from the chair of phonetics at the University of Edinburgh, and to make a beginning toward a systematic history of phonetics, one of the enduring interests of the distinguished scholar being saluted. As the editors note, this book is nothing like a complete or integrated history of the field, or even a first approximation to one; with nineteen authors from six countries and the inevitable diversity that multiple and independent authorship entails, this was not to be expected. It does, the editors say, include discussion of those matters that should be important components of any proper history. The papers are arranged into six "parts," three of which deal with the development of phonetic ideas (basic concepts, processes, voice quality, and voice dynamics), one on "national contributions," one on the achievements of individual scholars, and one on writing systems.

The contributions included in the first three parts of the book range over the following topics: feature classification (V. A. Fromkin and P. Ladefoged), the phonetics-phonology distinction of Kruszewski and the Kazan' School (K. H. Albow), the articulatory versus acoustic-auditory description of vowels (J. C. Catford), western traditions in the description of nasals (J. A. Kemp), early experimental studies of coarticulation (W. J. Hardcastle), consonantal rounding in British English (G. Brown), and the auditory analysis of voice quality (J. Laver) and prosody (M. Sumera). Of these papers, all but three are of interest largely as intended, that is, as history. The essays by Catford, Hardcastle, and Brown, while historically-oriented, might also engage the attention of an ahistorically-minded contemporary speech researcher. Catford makes a spirited yet judicious defense of the traditional vowel-height model of vowel classification developed by A. M. Bell and elaborated by Henry Sweet and Daniel Jones. Hardcastle's paper reminds us that the phenomenon of coarticulation, currently very much under scrutiny, is by no means a recent discovery, and that the view of speech as a

---

*Also Language in Society, 1983, 12, 369-371.
+Also University of Pennsylvania.
Acknowledgment. The preparation of this review was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-76 (1983)]

sequence of static positions or sounds linked by glides has been explicitly recognized as fallacious for at least a century. Brown discusses the relatively neglected feature of lip rounding as a property of British English consonants, casts doubt on historical inferences based on the absence of reference to it in earlier descriptions of the language, and suggests that the so-called rounded vowels are less reliably marked by rounding than are certain of the consonants.

Two papers deal with the contributions of individual phoneticians. One is by R. Thelwell, who describes the career of a relative, John Thelwell, a speech therapist and lecturer on "elocutionary science" in London during the late eighteenth and early nineteenth centuries. The other is a brief autobiographical sketch by K. L. Pike that describes the path he followed in his development as a phonetician; naturally enough, it does not begin to do justice to the contributions that make him the most accomplished phonetician among the American linguists of this and the last four decades.

W. S. Allen, who contributed significantly to our historical perspective with his study of the phonetics of ancient India, here gives an account of the phonetic thought of the ancient Greeks, and presents evidence to support his view that in phonetic analysis they were not up to the Indians, even if they elaborated an alphabet that he judges to be more useful a tool for phonetic analysis than the Devanagari. Two papers, by M. A. K. Halliday and by N-C. T. Chang, are densely packed with information on the development of phonetic and phonological thought in China. Both papers are rich in leads for those interested in pursuing the relation between phonetic theory and orthography in the Chinese context. Developments in phonetics in nineteenth-century Germany are, of course, much nearer home (hardly any more "national" than those in Britain), and K. Kohler goes well beyond a historical accounting to conduct a forceful polemic against what he deplores as the unfortunate separation of "linguistic phonetics" from phonetics in its physical, physiological, and psychological aspects, a development for which he holds Sievers mainly responsible. He argues vigorously against the separation of phonology from phonetics, which he terms an "ominous schism," and instead, champions the idea of an independent discipline of "speech science" freed of any "unfortunate" dependence on linguistics and open to all the disciplines that have something to contribute to the study of speech in all its aspects.

The last four papers are concerned with writing systems that can be said to represent, more or less imperfectly, the phonetic properties of speech. Two of them are devoted to the history of efforts to fit systems borrowed from one linguistic setting to another rather different one. A revision of a study by Abercrombie himself describes attempts over the past four centuries to repair the perceived deficiencies of the Latin (or Roman) alphabet as a vehicle for English, while J. Maw deals with the refashioning by scholars and others of the Arabic and Latin orthographies as devices for representing Swahili. Both papers go into detail in recording the many attepts at script "reform," but we can only infer from the extent to which the various changes proposed gained general acceptance just how widespread was dissatisfaction with the orthographical status quo. A paper by J. Kelly and one by M. K. C. MacMahon recount the history of the development of the various "shorthands" invented and promoted by a number of phoneticians during the late eighteenth and early nineteenth centuries: A. J. Ellis, I. Pitman, A. M. Bell, and Henry Sweet. These systems were, for the most part, intended to serve as auxiliaries to the standard orthography, for scientific and secretarial purposes. The modern

reader may be surprised to learn how large a role phoneticians played in a matter that aroused a good deal of public interest and contention at the time; the recent popularity of Shaw's Pygmalion brought about no revival of interest in the means by which Professor Higgins captured on paper the details of Liza's speech patterns.

A bibliography of David Abercrombie's published works, prepared by Elizabeth Uldall, a list of the names of persons and institutions whose subscriptions aided its publication, and indices of personal names and subjects complete the book. The editors and publisher are to be congratulated for the aesthetically pleasing format and remarkably error-free printing of an extremely demanding text.

Towards a history of phonetics succeeds in making a case for the serious study of the development of ideas about the nature of speech as an important aspect of the history of linguistics. Not that anyone would contend that the history of phonetic thought is unworthy of scholarly attention, but the contributors to this volume demonstrate that a wealth of readily accessible materials awaits the historian who can organize into a lucid picture men's opinions, different over time and place and cultures, regarding the nature of that uniquely human product, the act of speech.

II. PUBLICATIONS
III. APPENDIX

PUBLICATIONS

Bentin, S., & Feinsod, M. (1983). Hemispheric asymmetry for word perception: Biobehavioral and ERP evidence. Psychophysiology, 20, 489-497.

Bentin, S., Sahar, A., & Moscovitch, M. (in press). Intermanual transfer in patients with lesions in the trunk of the corpus callosum. Neuropsychologia.

Borden, G. J. (1983). Initiation versus execution time during manual and oral counting by stutterers. Journal of Speech and Hearing Research, 26, 389-396.

Crowder, R. G., & Repp, B. H. (in press). Single formant contrast in vowel identification. Perception & Psychophysics.

Goldstein, L. M. (1983). Vowel shifts and articulatory-acoustic relations. In A. Cohen & M. P. R. van den Broeke (Eds.), Abstracts of the Tenth International Congress of Phonetic Sciences (pp. 267-273). Dordrecht, The Netherlands: Foris Publications.

Hanson, V. L., Shankweiler, D., & Fischer, F. W. (1983). Determinants of spelling ability in deaf and hearing adults: Access to linguistic structure. Cognition, 14, 323-344.

Harris, K. S., & Bell-Berti, F. (1984). On consonants and syllable boundaries. In L. J. Raphael, C. B. Raphael, & M. R. Valdovinos (Eds.), Language and cognition (pp. 89-95). New York: Plenum Press.

Hoffman, P. R., Daniloff, R. G., Alfonso, P. J., & Schuckers, G. H. (in press). Multiple phoneme misarticulating children's perception and production of voice onset time. Perceptual and Motor Skills.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.

McGarr, N. S., & Gelfer, C. E. (1983). Simultaneous measurements of vowels produced by a hearing-impaired speaker. Language and Speech, 26, 233-246.

Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Communication, 2, 341-362.

Repp, B. H. (1984). Against a role of "chirp" identification in duplex perception. Perception & Psychophysics, 35, 89-93.

Repp, B. H. (in press). The role of release bursts in the perception of [s]-stop clusters. Journal of the Acoustical Society of America.

Tartter, V. C., Kat, D., Samuel, A. G., & Repp, B. H. (1983). Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions. Journal of the Acoustical Society of America, 74, 715-725.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. Perception & Psychophysics, 35, 49-64.

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	AD A129713	ED 229-816
SR-74/75	April - September 1983	AD A136416	ED 236-753

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Haskins Laboratories
270 Crown Street
New Haven, CT 06511

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Haskins Laboratories Status Report on Speech Research, SR-76, October - December, 1983

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)
Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE
December, 1983

7a. TOTAL NO. OF PAGES
262

7b. NO. OF REFS
370

8a. CONTRACT OR GRANT NO.

HD-01994	NS 13870
HD-16591	NS 13617
NO 1-HD-1-2420	NS 18010
RR-0559€	NS 07237
BNS-8111470	NS 07196
N00014-83-C-0083	

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-76 (1983)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
See No. 8

13. ABSTRACT

This report (1 October-31 December) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

-Obituary--Dennis Butler Fry
-Skilled actions: A task dynamic approach
-Speculations on the control of fundamental frequency declination
-Selective effects of masking on speech and nonspeech in the duplex perception paradigm
-Vowels in consonantal context are perceived more linguistically than isolated vowels: Evidence from an individual differences scaling study
-Children's perception of [s] and [ʃ]: The relation between articulation and perceptual adjustment for coarticulatory effects
-Trading relations among acoustic cues in speech perception: Speech-specific but not special
-The role of release bursts in the perception of [s]-stop clusters
-A perceptual analog of change in progress in Welsh
-Single formant contrast in vowel identification
-Integration of melody and text in memory for songs
-The equation of information and meaning from the perspectives of situation semantics and Gibson's ecological realism
-A comment on the equating of information with symbol strings
-An ecological approach to perception and action
-Mapping speech: More analysis, less synthesis, please
-Review (Towards a history of phonetics)

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception:						
masking, speech, nonspeech, duplex perception						
vowels in consonantal context, isolated vowels						
individual differences, scaling						
fricatives, coarticulation, adjustments						
trading relations, cues, stop clusters, release bursts						
language change, Welsh						
formant contrast, vowel identification						
mapping, analysis, synthesis						
phonetics, history						
Speech Articulation:						
declination, fundamental frequency, control						
Reading:						
songs, text, melody, integration						
information, meaning, semantics						
ecological realism, Gibson, symbol strings						
Motor Control:						
skill, action, task dynamics						
perception, action, ecological approach						
Obituary:						
Fry, Dennis Butler						